# RAPID DETECTION AND IDENTIFICATION OF PATHOGENS

## FIELD OF THE INVENTION

The present invention relates to methods and compositions for treating nucleic acid, and in particular, methods and compositions for detection and characterization of nucleic acid sequences and sequence changes.

## BACKGROUND OF THE INVENTION

The detection and characterization of specific nucleic acid sequences and sequence changes have been utilized to detect the presence of viral or bacterial nucleic acid sequences indicative of an infection, the presence of variants or alleles of mammalian genes associated with disease and cancers, and the identification of the source of nucleic acids found in forensic samples, as well as in paternity determinations.

Various methods are known in the art which may be used to detect and characterize specific nucleic acid sequences and sequence changes. Nonetheless, as nucleic acid sequence data of the human genome, as well as the genomes of pathogenic organisms accumulates, the demand for fast, reliable, cost-effective and user-friendly tests for specific sequences continues to grow. Importantly, these tests must be able to create a detectable signal from a very low copy number of the sequence of interest. The following discussion examines three levels of nucleic acid detection currently in use: I. Signal Amplification Technology for detection of rare sequences; II. Direct Detection Technology for detection of higher copy number sequences; and III. Detection of Unknown Sequence Changes for rapid screening of sequence changes anywhere within a defined DNA fragment.

### I. Signal Amplification Technology Methods For Amplification

The "Polymerase Chain Reaction" (PCR) comprises the first generation of methods for nucleic acid amplification. However, several other methods have been

developed that employ the same basis of specificity, but create signal by different amplification mechanisms. These methods include the "Ligase Chain Reaction" (LCR), "Self-Sustained Synthetic Reaction" (3SR/NASBA), and "Qβ-Replicase" (Qβ).

**Polymerase Chain Reaction** (PCR)

The polymerase chain reaction (PCR), as described in U.S. Patent Nos. 4,683,195 and 4,683,202 to Mullis and Mullis *et al.*, describe a method for increasing the concentration of a segment of target sequence in a mixture of genomic DNA without cloning or purification. This technology provides one approach to the problems of low target sequence concentration. PCR can be used to directly increase the concentration of the target to an easily detectable level. This process for amplifying the target sequence involves introducing a molar excess of two oligonucleotide primers which are complementary to their respective strands of the double-stranded target sequence to the DNA mixture containing the desired target sequence. The mixture is denatured and then allowed to hybridize. Following hybridization, the primers are extended with polymerase so as to form complementary strands. The steps of denaturation, hybridization, and polymerase extension can be repeated as often as needed, in order to obtain relatively high concentrations of a segment of the desired target sequence.

The length of the segment of the desired target sequence is determined by the relative positions of the primers with respect to each other, and, therefore, this length is a controllable parameter. Because the desired segments of the target sequence become the dominant sequences (in terms of concentration) in the mixture, they are said to be "PCR-amplified."

**Ligase Chain Reaction** (LCR or LAR)

The ligase chain reaction (LCR; sometimes referred to as "Ligase Amplification Reaction" (LAR) described by Barany, Proc. Natl. Acad. Sci., 88:189 (1991); Barany, PCR Methods and Applic., 1:5 (1991); and Wu and Wallace, Genomics 4:560 (1989) has developed into a well-recognized alternative method for amplifying nucleic acids.

In LCR, four oligonucleotides, two adjacent oligonucleotides which uniquely hybridize to one strand of target DNA, and a complementary set of adjacent oligonucleotides, which hybridize to the opposite strand are mixed and DNA ligase is added to the mixture. Provided that there is complete complementarity at the junction, ligase will covalently link each set of hybridized molecules. Importantly, in LCR, two probes are ligated together only when they base-pair with sequences in the target sample, without gaps or mismatches. Repeated cycles of denaturation, hybridization and ligation amplify a short segment of DNA. LCR has also been used in combination with PCR to achieve enhanced detection of single-base changes. Segev, PCT Public. No. W09001069 A1 (1990). However, because the four oligonucleotides used in this assay can pair to form two short ligatable fragments, there is the potential for the generation of target-independent background signal. The use of LCR for mutant screening is limited to the examination of specific nucleic acid positions.

**Self-Sustained Synthetic Reaction** (3SR/NASBA)

The self-sustained sequence replication reaction (3SR) (Guatelli *et al.*, Proc. Natl. Acad. Sci., 87:1874-1878 [1990], with an erratum at Proc. Natl. Acad. Sci., 87:7797 [1990]) is a transcription-based *in vitro* amplification system (Kwok *et al.*, Proc. Natl. Acad. Sci., 86:1173-1177 [1989]) that can exponentially amplify RNA sequences at a uniform temperature. The amplified RNA can then be utilized for mutation detection (Fahy *et al.*, PCR Meth. Appl., 1:25-33 [1991]). In this method, an oligonucleotide primer is used to add a phage RNA polymerase promoter to the 5' end of the sequence of interest. In a cocktail of enzymes and substrates that includes a second primer, reverse transcriptase, RNase H, RNA polymerase and ribo-and deoxyribonucleoside triphosphates, the target sequence undergoes repeated rounds of transcription, cDNA synthesis and second-strand synthesis to amplify the area of interest. The use of 3SR to detect mutations is kinetically limited to screening small segments of DNA (*e.g.*, 200-300 base pairs).

**Q-Beta (Qβ) Replicase**

In this method, a probe which recognizes the sequence of interest is attached to the replicatable RNA template for Qβ replicase. A previously identified major problem with false positives resulting from the replication of unhybridized probes has been addressed through use of a sequence-specific ligation step. However, available thermostable DNA ligases are not effective on this RNA substrate, so the ligation must be performed by T4 DNA ligase at low temperatures (37°C). This prevents the use of high temperature as a means of achieving specificity as in the LCR, the ligation event can be used to detect a mutation at the junction site, but not elsewhere.

Table 1 below, lists some of the features desirable for systems useful in sensitive nucleic acid diagnostics, and summarizes the abilities of each of the major amplification methods (*See also*, Landgren, Trends in Genetics 9:199 [1993]).

A successful diagnostic method must be very specific. A straight-forward method of controlling the specificity of nucleic acid hybridization is by controlling the temperature of the reaction. While the 3SR/NASBA, and Qβ systems are all able to generate a large quantity of signal, one or more of the enzymes involved in each cannot be used at high temperature (*i.e.*, >55°C). Therefore the reaction temperatures cannot be raised to prevent non-specific hybridization of the probes. If probes are shortened in order to make them melt more easily at low temperatures, the likelihood of having more than one perfect match in a complex genome increases. For these reasons, PCR and LCR currently dominate the research field in detection technologies.

TABLE 1

| FEATURE | METHOD: | | | | |
|---|---|---|---|---|---|
| | PCR | LCR | PCR & LCR | 3SR NASBA | Qβ |
| Amplifies Target | + | + | + | + | |
| Recognition of Independent Sequences Required | + | + | + | + | + |
| Performed at High Temp. | + | + | | | |
| Operates at Fixed Temp. | | | | + | + |
| Exponential Amplification | + | + | + | + | + |
| Generic Signal Generation | | | | | + |
| Easily Automatable | | | | | |

The basis of the amplification procedure in the PCR and LCR is the fact that the products of one cycle become usable templates in all subsequent cycles, consequently doubling the population with each cycle. The final yield of any such doubling system can be expressed as: $(1+X)^n = y$, where "X" is the mean efficiency (percent copied in each cycle), "n" is the number of cycles, and "y" is the overall efficiency, or yield of the reaction (Mullis, PCR Methods Applic., 1:1 [1991]). If every copy of a target DNA is utilized as a template in every cycle of a polymerase chain reaction, then the mean efficiency is 100%. If 20 cycles of PCR are performed, then the yield will be $2^{20}$, or 1,048,576 copies of the starting material. If the reaction conditions reduce the mean efficiency to 85%, then the yield in those 20 cycles will be only $1.85^{20}$, or 220,513 copies of the starting material. In other words, a PCR running

at 85% efficiency will yield only 21% as much final product, compared to a reaction running at 100% efficiency. A reaction that is reduced to 50% mean efficiency will yield less than 1% of the possible product.

In practice, routine polymerase chain reactions rarely achieve the theoretical maximum yield, and PCRs are usually run for more than 20 cycles to compensate for the lower yield. At 50% mean efficiency, it would take 34 cycles to achieve the million-fold amplification theoretically possible in 20, and at lower efficiencies, the number of cycles required becomes prohibitive. In addition, any background products that amplify with a better mean efficiency than the intended target will become the dominant products.

Also, many variables can influence the mean efficiency of PCR, including target DNA length and secondary structure, primer length and design, primer and dNTP concentrations, and buffer composition, to name but a few. Contamination of the reaction with exogenous DNA (*e.g.*, DNA spilled onto lab surfaces) or cross-contamination is also a major consideration. Reaction conditions must be carefully optimized for each different primer pair and target sequence, and the process can take days, even for an experienced investigator. The laboriousness of this process, including numerous technical considerations and other factors, presents a significant drawback to using PCR in the clinical setting. Indeed, PCR has yet to penetrate the clinical market in a significant way. The same concerns arise with LCR, as LCR must also be optimized to use different oligonucleotide sequences for each target sequence. In addition, both methods require expensive equipment, capable of precise temperature cycling.

Many applications of nucleic acid detection technologies, such as in studies of allelic variation, involve not only detection of a specific sequence in a complex background, but also the discrimination between sequences with few, or single, nucleotide differences. One method for the detection of allele-specific variants by PCR is based upon the fact that it is difficult for *Taq* polymerase to synthesize a DNA strand when there is a mismatch between the template strand and the 3' end of the primer. An allele-specific variant may be detected by the use of a primer that is

perfectly matched with only one of the possible alleles; the mismatch to the other allele acts to prevent the extension of the primer, thereby preventing the amplification of that sequence. This method has a substantial limitation in that the base composition of the mismatch influences the ability to prevent extension across the mismatch, and certain mismatches do not prevent extension or have only a minimal effect (Kwok *et al.*, Nucl. Acids Res., 18:999 [1990]).)

A similar 3'-mismatch strategy is used with greater effect to prevent ligation in the LCR (Barany, PCR Meth. Applic., 1:5 [1991]). Any mismatch effectively blocks the action of the thermostable ligase, but LCR still has the drawback of target-independent background ligation products initiating the amplification. Moreover, the combination of PCR with subsequent LCR to identify the nucleotides at individual positions is also a clearly cumbersome proposition for the clinical laboratory.

## II. Direct Detection Technology

When a sufficient amount of a nucleic acid to be detected is available, there are advantages to detecting that sequence directly, instead of making more copies of that target, (*e.g.*, as in PCR and LCR). Most notably, a method that does not amplify the signal exponentially is more amenable to quantitative analysis. Even if the signal is enhanced by attaching multiple dyes to a single oligonucleotide, the correlation between the final signal intensity and amount of target is direct. Such a system has an additional advantage that the products of the reaction will not themselves promote further reaction, so contamination of lab surfaces by the products is not as much of a concern. Traditional methods of direct detection including Northern and Southern blotting and RNase protection assays usually require the use of radioactivity and are not amenable to automation. Recently devised techniques have sought to eliminate the use of radioactivity and/or improve the sensitivity in automatable formats. Two examples are the "Cycling Probe Reaction" (CPR), and "Branched DNA" (bDNA)

The cycling probe reaction (CPR) (Duck *et al.*, BioTech., 9:142 [1990]), uses a long chimeric oligonucleotide in which a central portion is made of RNA while the two termini are made of DNA. Hybridization of the probe to a target DNA and exposure to a thermostable RNase H causes the RNA portion to be digested. This destabilizes the remaining DNA portions of the duplex, releasing the remainder of the probe from the target DNA and allowing another probe molecule to repeat the process. The signal, in the form of cleaved probe molecules, accumulates at a linear rate. While the repeating process increases the signal, the RNA portion of the oligonucleotide is vulnerable to RNases that may carried through sample preparation.

Branched DNA (bDNA), described by Urdea *et al.*, Gene 61:253-264 (1987), involves oligonucleotides with branched structures that allow each individual oligonucleotide to carry 35 to 40 labels (*e.g.*, alkaline phosphatase enzymes). While this enhances the signal from a hybridization event, signal from non-specific binding is similarly increased.

## III. Detection Of Unknown Sequence Changes

The demand for tests which allow the detection of specific nucleic acid sequences and sequence changes is growing rapidly in clinical diagnostics. As nucleic acid sequence data for genes from humans and pathogenic organisms accumulates, the demand for fast, cost-effective, and easy-to-use tests for as yet unknown mutations within specific sequences is rapidly increasing.

A handful of methods have been devised to scan nucleic acid segments for mutations. One option is to determine the entire gene sequence of each test sample (*e.g.*, a bacterial isolate). For sequences under approximately 600 nucleotides, this may be accomplished using amplified material (*e.g.*, PCR reaction products). This avoids the time and expense associated with cloning the segment of interest. However, specialized equipment and highly trained personnel are required, and the method is too labor-intense and expensive to be practical and effective in the clinical setting.

In view of the difficulties associated with sequencing, a given segment of nucleic acid may be characterized on several other levels. At the lowest resolution, the size of the molecule can be determined by electrophoresis by comparison to a known standard run on the same gel. A more detailed picture of the molecule may be achieved by cleavage with combinations of restriction enzymes prior to electrophoresis, to allow construction of an ordered map. The presence of specific sequences within the fragment can be detected by hybridization of a labeled probe, or the precise nucleotide sequence can be determined by partial chemical degradation or by primer extension in the presence of chain-terminating nucleotide analogs.

For detection of single-base differences between like sequences, the requirements of the analysis are often at the highest level of resolution. For cases in which the position of the nucleotide in question is known in advance, several methods have been developed for examining single base changes without direct sequencing. For example, if a mutation of interest happens to fall within a restriction recognition sequence, a change in the pattern of digestion can be used as a diagnostic tool (*e.g.*, restriction fragment length polymorphism [RFLP] analysis).

Single point mutations have been also detected by the creation or destruction of RFLPs. Mutations are detected and localized by the presence and size of the RNA fragments generated by cleavage at the mismatches. Single nucleotide mismatches in DNA heteroduplexes are also recognized and cleaved by some chemicals, providing an alternative strategy to detect single base substitutions, generically named the "Mismatch Chemical Cleavage" (MCC) (Gogos *et al.*, Nucl. Acids Res., 18:6807-6817 [1990]). However, this method requires the use of osmium tetroxide and piperidine, two highly noxious chemicals which are not suited for use in a clinical laboratory.

RFLP analysis suffers from low sensitivity and requires a large amount of sample. When RFLP analysis is used for the detection of point mutations, it is, by its nature, limited to the detection of only those single base changes which fall within a restriction sequence of a known restriction endonuclease. Moreover, the majority of the available enzymes have 4 to 6 base-pair recognition sequences, and cleave too frequently for many large-scale DNA manipulations (Eckstein and Lilley (eds.),

*Nucleic Acids and Molecular Biology*, vol. 2, Springer-Verlag, Heidelberg [1988]). Thus, it is applicable only in a small fraction of cases, as most mutations do not fall within such sites.

A handful of rare-cutting restriction enzymes with 8 base-pair specificities have been isolated and these are widely used in genetic mapping, but these enzymes are few in number, are limited to the recognition of G+C-rich sequences, and cleave at sites that tend to be highly clustered (Barlow and Lehrach, Trends Genet., 3:167 [1987]). Recently, endonucleases encoded by group I introns have been discovered that might have greater than 12 base-pair specificity (Perlman and Butow, Science 246:1106 [1989]), but again, these are few in number.

If the change is not in a recognition sequence, then allele-specific oligonucleotides (ASOs), can be designed to hybridize in proximity to the unknown nucleotide, such that a primer extension or ligation event can be used as the indicator of a match or a mis-match. Hybridization with radioactively labeled allelic specific oligonucleotides (ASO) also has been applied to the detection of specific point mutations (Conner *et al.*, Proc. Natl. Acad. Sci., 80:278-282 [1983]). The method is based on the differences in the melting temperature of short DNA fragments differing by a single nucleotide. Stringent hybridization and washing conditions can differentiate between mutant and wild-type alleles. The ASO approach applied to PCR products also has been extensively utilized by various researchers to detect and characterize point mutations in *ras* genes (Vogelstein *et al.*, N. Eng. J. Med., 319:525-532 [1988]; and Farr *et al.*, Proc. Natl. Acad. Sci., 85:1629-1633 [1988]), and *gsp/gip* oncogenes (Lyons *et al.*, Science 249:655-659 [1990]). Because of the presence of various nucleotide changes in multiple positions, the ASO method requires the use of many oligonucleotides to cover all possible oncogenic mutations.

With either of the techniques described above (*i.e.*, RFLP and ASO), the precise location of the suspected mutation must be known in advance of the test. That is to say, they are inapplicable when one needs to detect the presence of a mutation of an unknown character and position within a gene or sequence of interest.

Two other methods rely on detecting changes in electrophoretic mobility in response to minor sequence changes. One of these methods, termed "Denaturing Gradient Gel Electrophoresis" (DGGE) is based on the observation that slightly different sequences will display different patterns of local melting when electrophoretically resolved on a gradient gel. In this manner, variants can be distinguished, as differences in melting properties of homoduplexes versus heteroduplexes differing in a single nucleotide can detect the presence of mutations in the target sequences because of the corresponding changes in their electrophoretic mobilities. The fragments to be analyzed, usually PCR products, are "clamped" at one end by a long stretch of G-C base pairs (30-80) to allow complete denaturation of the sequence of interest without complete dissociation of the strands. The attachment of a GC "clamp" to the DNA fragments increases the fraction of mutations that can be recognized by DGGE (Abrams *et al.*, Genomics 7:463-475 [1990]). Attaching a GC clamp to one primer is critical to ensure that the amplified sequence has a low dissociation temperature (Sheffield *et al.*, Proc. Natl. Acad. Sci., 86:232-236 [1989]; and Lerman and Silverstein, Meth. Enzymol., 155:482-501 [1987]). Modifications of the technique have been developed, using temperature gradients (Wartell *et al.*, Nucl. Acids Res., 18:2699-2701 [1990]), and the method can be also applied to RNA:RNA duplexes (Smith *et al.*, Genomics 3:217-223 [1988]).

Limitations on the utility of DGGE include the requirement that the denaturing conditions must be optimized for each type of DNA to be tested. Furthermore, the method requires specialized equipment to prepare the gels and maintain the needed high temperatures during electrophoresis. The expense associated with the synthesis of the clamping tail on one oligonucleotide for each sequence to be tested is also a major consideration. In addition, long running times are required for DGGE. The long running time of DGGE was shortened in a modification of DGGE called constant denaturant gel electrophoresis (CDGE) (Borrensen *et al.*, Proc. Natl. Acad. Sci. USA 88:8405 [1991]). CDGE requires that gels be performed under different denaturant conditions in order to reach high efficiency for the detection of unknown mutations.

An technique analogous to DGGE, termed temperature gradient gel electrophoresis (TGGE), uses a thermal gradient rather than a chemcial denaturant gradient (Scholz, *et al.*, Hum. Mol. Genet. 2:2155 [1993]). TGGE requires the use of specialized equipment which can generate a temperature gradient perpendicularly oriented relative to the electrical field. TGGE can detect mutations in relatively small fragments of DNA therefore scanning of large gene segments requires the use of multiple PCR products prior to running the gel.

Another common method, called "Single-Strand Conformation Polymorphism" (SSCP) was developed by Hayashi, Sekya and colleagues (reviewed by Hayashi, PCR Meth. Appl., 1:34-38, [1991]) and is based on the observation that single strands of nucleic acid can take on characteristic conformations in non-denaturing conditions, and these conformations influence electrophoretic mobility. The complementary strands assume sufficiently different structures that one strand may be resolved from the other. Changes in sequences within the fragment will also change the conformation, consequently altering the mobility and allowing this to be used as an assay for sequence variations (Orita, *et al.*, Genomics 5:874-879, [1989]).

The SSCP process involves denaturing a DNA segment (*e.g.*, a PCR product) that is labelled on both strands, followed by slow electrophoretic separation on a non-denaturing polyacrylamide gel, so that intra-molecular interactions can form and not be disturbed during the run. This technique is extremely sensitive to variations in gel composition and temperature. A serious limitation of this method is the relative difficulty encountered in comparing data generated in different laboratories, under apparently similar conditions.

The dideoxy fingerprinting (ddF) is another technique developed to scan genes for the presence of unknown mutations (Liu and Sommer, PCR Methods Appli., 4:97 [1994]). The ddF technique combines components of Sanger dideoxy sequencing with SSCP. A dideoxy sequencing reaction is performed using one dideoxy terminator and then the reaction products are electrophoresised on nondenaturing polyacrylamide gels to detect alterations in mobility of the termination segments as in SSCP analysis. While ddF is an improvement over SSCP in terms of increased sensitivity, ddF

requires the use of expensive dideoxynucleotides and this technique is still limited to the analysis of fragments of the size suitable for SSCP (*i.e.*, fragments of 200-300 bases for optimal detection of mutations).

In addition to the above limitations, all of these methods are limited as to the size of the nucleic acid fragment that can be analyzed. For the direct sequencing approach, sequences of greater than 600 base pairs require cloning, with the consequent delays and expense of either deletion sub-cloning or primer walking, in order to cover the entire fragment. SSCP and DGGE have even more severe size limitations. Because of reduced sensitivity to sequence changes, these methods are not considered suitable for larger fragments. Although SSCP is reportedly able to detect 90% of single-base substitutions within a 200 base-pair fragment, the detection drops to less than 50% for 400 base pair fragments. Similarly, the sensitivity of DGGE decreases as the length of the fragment reaches 500 base-pairs. The ddF technique, as a combination of direct sequencing and SSCP, is also limited by the relatively small size of the DNA that can be screened.

Clearly, there remains a need for a method that is less sensitive to size so that entire genes, rather than gene fragments, may be analyzed. Such a tool must also be robust, so that data from different labs, generated by researchers of diverse backgrounds and skills will be comparable. Ideally, such a method would be compatible with "multiplexing," (*i.e.*, the simultaneous analysis of several molecules or genes in a single reaction or gel lane, usually resolved from each other by differential labelling or probing). Such an analytical procedure would facilitate the use of internal standards for subsequent analysis and data comparison, and increase the productivity of personnel and equipment. The ideal method would also be easily automatable.

## SUMMARY OF THE INVENTION

The present invention relates to methods and compositions for treating nucleic acid, and in particular, methods and compositions for detection and characterization of nucleic acid sequences and sequence changes in microbial gene sequences. The present invention provides means for cleaving a nucleic acid cleavage structure in a

site-specific manner. In one embodiment, the means for cleaving is an enzyme capable of cleaving cleavage structures on a nucleic acid substrate, forming the basis of a novel method of detection of specific nucleic acid sequences. The present invention contemplates use of the novel detection method for, among other uses, clinical diagnostic purposes, including but not limited to detection and identification of pathogenic organisms.

In one embodiment, the present invention contemplates a DNA sequence encoding a DNA polymerase altered in sequence (*i.e.*, a "mutant" DNA polymerase) relative to the native sequence such that it exhibits altered DNA synthetic activity from that of the native (*i.e.*, "wild type") DNA polymerase. With regard to the polymerase, a complete absence of synthesis is not required; it is desired that cleavage reactions occur in the absence of polymerase activity at a level where it interferes with the method. It is preferred that the encoded DNA polymerase is altered such that it exhibits reduced synthetic activity from that of the native DNA polymerase. In this manner, the enzymes of the invention are nucleases and are capable of cleaving nucleic acids in a structure-specific manner. Importantly, the nucleases of the present invention are capable of cleaving cleavage structures to create discrete cleavage products.

The present invention contemplates nucleases from a variety of sources, including nucleases that are thermostable. Thermostable nucleases are contemplated as particularly useful, as they are capable of operating at temperatures where nucleic acid hybridization is extremely specific, allowing for allele-specific detection (including single-base mismatches). In one embodiment, the thermostable 5' nucleases are selected from the group consisting of altered polymerases derived from the native polymerases of various *Thermus* species, including, but not limited to *Thermus aquaticus*, *Thermus flavus* and *Thermus thermophilus*.

The present invention utilizes such enzymes in methods for detection and characterization of nucleic acid sequences and sequence changes. The present invention relates to means for cleaving a nucleic acid cleavage structure in a site-

specific manner. Nuclease activity is used to screen for known and unknown mutations, including single base changes, in nucleic acids.

In one embodiment, the present invention contemplates a process or method for identifying strains of microorganisms comprising the steps of providing a cleavage means and a nucleic acid substrate containing sequences derived from one or more microorganism; treating the nucleic acid substrate under conditions such that the substrate forms one or more cleavage structures; and reacting the cleavage means with the cleavage structures so that one or more cleavage products are produced. In one embodiment of this invention, the cleavage means is an enzyme. In one preferred embodiment, the enzyme is a nuclease. In an alternative preferred embodiment, the nuclease is selected from the group consisting of Cleavase™ BN, *Thermus aquaticus* DNA polymerase, *Thermus thermophilus* DNA polymerase, *Escherichia coli* Exo III, and the *Saccharomyces cerevisiae* Rad1/Rad10 complex. It is also contemplated that the enzyme may have a portion of its amino acid sequence that is homologous to a portion of the amino acid sequence of a thermostable DNA polymerase derived from a eubacterial thermophile, the latter being selected from the group consisting of *Thermus aquaticus*, *Thermus flavus* and *Thermus thermophilus*.

It is contemplated that the nucleic acid substrate comprise a nucleotide analog, including but not limited to the group comprising 7-deaza-dATP, 7-deaza-dGTP and dUTP. In one embodiment, the nucleic acid substrate is substantially single-stranded. It is not intended that the nucleic acid substrate be limited to any particular form, indeed, it is contemplated that the nucleic acid substrate is single stranded or double-stranded RNA or DNA.

In one embodiment of the present invention, the treating step comprises rendering double-stranded nucleic acid substantially single-stranded, and exposing the single-stranded nucleic acid to conditions such that the single-stranded nucleic acid assumes a secondary or characteristic folded structure. In one preferred embodiment, double-stranded nucleic acid is rendered substantially single-stranded by increased temperature.

In an alternative embodiment, the method of the present invention further comprises the step of detecting one or more cleavage products.

It is contemplated that the microorganism(s) of the present invention be selected from a variety of microorganisms. It is not intended that the present invention be limited to any particular type of microorganism. Rather, it is intended that the present invention be used with organisms including, but not limited to, bacteria, fungi, protozoa, ciliates, and viruses. It is not intended that the microorganisms be limited to a particular genus, species, strain, or serotype. Indeed, it is contemplated that the bacteria be selected from the group including, but not limited to members of the genera *Campylobacter, Escherichia, Mycobacterium, Salmonella, Shigella,* and *Staphylococcus*. In one preferred embodiment, the microorganism(s) comprise strains of multi-drug resistant *Mycobacterium tuberculosis*. It is also contemplated that the present invention be used with viruses, including but not limited to hepatitis C virus and simian immunodeficiency virus.

Another embodiment of the present invention contemplates a method for detecting and identifying strains of microorganisms, comprising the steps of extracting nucleic acid from a sample suspected of containing one or more microorganisms; and contacting the extracted nucleic acid with a cleavage means under conditions such that the extracted nucleic acid forms one or more secondary structures, and the cleavage means cleaves the secondary structures to produce one or more cleavage products.

In one embodiment, the method further comprises the step of separating the cleavage products. In yet another embodiment, the method further comprises the step of detecting the cleavage products.

In one preferred embodiment, the present invention further comprises comparing the detected cleavage products generated from cleavage of the extracted nucleic acid isolated from the sample with separated cleavage products generated by cleavage of nucleic acids derived from one or more reference microorganisms. In such a case the sequence of the nucleic acids from one or more reference microorganisms may be related but different (*e.g.*, a wild type control for a mutant sequence or a known or previously characterized mutant sequence).

In an alternative preferred embodiment, the present invention further comprises the step of isolating a polymorphic locus from the extracted nucleic acid after the extraction step, so as to generate a nucleic acid substrate, wherein the substrate is contacted with the cleavage means. In one embodiment, the isolation of a polymorphic locus is accomplished by polymerase chain reaction amplification. In an alternate embodiment, the polymerase chain reaction is conducted in the presence of a nucleotide analog, including but not limited to the group comprising 7-deaza-dATP, 7-deaza-dGTP and dUTP. It is contemplated that the polymerase chain reaction amplification will employ oligonucleotide primers matching or complementary to consensus gene sequences derived from the polymorphic locus. In one embodiment, the polymorphic locus comprises a ribosomal RNA gene. In a particularly preferred embodiment, the ribosomal RNA gene is a 16S ribosomal RNA gene.

In one embodiment of this method, the cleavage means is an enzyme. In one preferred embodiment, the enzyme is a nuclease. In a particularly preferred embodiment, the nuclease is selected from the group including, but not limited to Cleavase™ BN, *Thermus aquaticus* DNA polymerase, *Thermus thermophilus* DNA polymerase, *Escherichia coli* Exo III, and the *Saccharomyces cerevisiae* Rad1/Rad10 complex. It is also contemplated that the enzyme may have a portion of its amino acid sequence that is homologous to a portion of the amino acid sequence of a thermostable DNA polymerase derived from a eubacterial thermophile, the latter being selected from the group consisting of *Thermus aquaticus*, *Thermus flavus* and *Thermus thermophilus*.

It is contemplated that the nucleic acid substrate of this method will comprise a nucleotide analog, including but not limited to the group comprising 7-deaza-dATP, 7-deaza-dGTP and dUTP. In one embodiment, the nucleic acid substrate is substantially single-stranded. It is not intended that the nucleic acid substrate be limited to any particular form, indeed, it is contemplated that the nucleic acid substrate is single stranded or double-stranded RNA or DNA.

In another embodiment of the present invention, the treating step of the method comprises rendering double-stranded nucleic acid substantially single-stranded, and

exposing the single-stranded nucleic acid to conditions such that the single-stranded nucleic acid has secondary structure. In one preferred embodiment, double-stranded nucleic acid is rendered substantially single-stranded by increased temperature.

It is contemplated that the microorganism(s) of the present invention be selected from a variety of microorganisms; it is not intended that the present invention be limited to any particular type of microorganism. Rather, it is intended that the present invention will be used with organisms including, but not limited to, bacteria, fungi, protozoa, ciliates, and viruses. It is not intended that the microorganisms be limited to a particular genus, species, strain, or serotype. Indeed, it is contemplated that the bacteria be selected from the group comprising, but not limited to members of the genera *Campylobacter, Escherichia, Mycobacterium, Salmonella, Shigella,* and *Staphylococcus*. In one preferred embodiment, the microorganism(s) comprise strains of multi-drug resistant *Mycobacterium tuberculosis*. It is also contemplated that the present invention be used with viruses, including but not limited to hepatitis C virus and simian immunodeficiency virus.

In yet another embodiment, the present invention contemplates a method for treating nucleic acid comprising an oligonucleotide containing microbial gene sequences, comprising providing a cleavage means in a solution containing manganese and nucleic acid substrate containing microbial gene sequences; treating the nucleic acid substrate with increased temperature such that the substrate is substantially single-stranded; reducing the temperature under conditions such that the single-stranded substrate forms one or more cleavage structures; reacting the cleavage means with the cleavage structures so that one or more cleavage products are produced; and detecting the one or more cleavage products produced by the method.

The present invention also contemplates a process for creating a record reference library of genetic fingerprints characteristic (*i.e.*, diagnostic) of one or more alleles of the various microorganisms, comprising the steps of providing a cleavage means and nucleic acid substrate derived from microbial gene sequences; contacting the nucleic acid substrate with a cleavage means under conditions such that the extracted nucleic acid forms one or more secondary structures and the cleavage means

cleaves the secondary structures, resulting in the generation of multiple cleavage products; separating the multiple cleavage products; and maintaining a testable record reference of the separated cleavage products.

By the term "genetic fingerprint" it is meant that changes in the sequence of the nucleic acid (*e.g.*, a deletion, insertion or a single point substitution) alter the structures formed, thus changing the banding pattern (*i.e.*, the "fingerprint" or "bar code") to reflect the difference in the sequence, allowing rapid detection and identification of variants.

The methods of the present invention allow for simultaneous analysis of both strands (*e.g.*, the sense and antisense strands) and are ideal for high-level multiplexing. The products produced are amenable to qualitative, quantitative and positional analysis. The methods may be automated and may be practiced in solution or in the solid phase (*e.g.*, on a solid support). The methods are powerful in that they allow for analysis of longer fragments of nucleic acid than current methodologies.

## DESCRIPTION OF THE DRAWINGS

Figure 1A provides a schematic of one embodiment of the detection method of the present invention.

Figure 1B provides a schematic of a second embodiment of the detection method of the present invention.

Figure 2 is a comparison of the nucleotide structure of the DNAP genes isolated from *Thermus aquaticus* (SEQ ID NO:1), *Thermus flavus* (SEQ ID NO:2) and *Thermus thermophilus* (SEQ ID NO:3); the consensus sequence (SEQ ID NO:7) is shown at the top of each row.

Figure 3 is a comparison of the amino acid sequence of the DNAP isolated from *Thermus aquaticus* (SEQ ID NO:4), *Thermus flavus* (SEQ ID NO:5), and *Thermus thermophilus*(SEQ ID NO:6); the consensus sequence (SEQ ID NO:8) is shown at the top of each row.

Figures 4A-G are a set of diagrams of wild-type and synthesis-deficient DNAP*Taq* genes.

Figure 5A depicts the wild-type *Thermus flavus* polymerase gene.

Figure 5B depicts a synthesis-deficient *Thermus flavus* polymerase gene.

Figure 6 depicts a structure which cannot be amplified using DNAP*Taq*.

Figure 7 is a ethidium bromide-stained gel demonstrating attempts to amplify a bifurcated duplex using either DNAP*Taq* or DNAPStf (Stoffel).

Figure 8 is an autoradiogram of a gel analyzing the cleavage of a bifurcated duplex by DNAP*Taq* and lack of cleavage by DNAPStf.

Figures 9A-B are a set of autoradiograms of gels analyzing cleavage or lack of cleavage upon addition of different reaction components and change of incubation temperature during attempts to cleave a bifurcated duplex with DNAP*Taq*.

Figures 10A-B are an autoradiogram displaying timed cleavage reactions, with and without primer.

Figures 11A-B are a set of autoradiograms of gels demonstrating attempts to cleave a bifurcated duplex (with and without primer) with various DNAPs.

Figures 12A shows the substrates and oligonucleotides used to test the specific cleavage of substrate DNAs targeted by pilot oligonucleotides.

Figure 12B shows an autoradiogram of a gel showing the results of cleavage reactions using the substrates and oligonucleotides shown Fig. 12A.

Figure 13A shows the substrate and oligonucleotide used to test the specific cleavage of a substrate RNA targeted by a pilot oligonucleotide.

Figure 13B shows an autoradiogram of a gel showing the results of a cleavage reaction using the substrate and oligonucleotide shown in Fig. 13A.

Figure 14 is a diagram of vector pTTQ18.

Figure 15 is a diagram of vector pET-3c.

Figure 16A-E depicts a set of molecules which are suitable substrates for cleavage by the 5' nuclease activity of DNAPs.

Figure 17 is an autoradiogram of a gel showing the results of a cleavage reaction run with synthesis-deficient DNAPs.

Figure 18 is an autoradiogram of a PEI chromatogram resolving the products of an assay for synthetic activity in synthesis-deficient DNAP*Taq* clones.

Figure 19A depicts the substrate molecule used to test the ability of synthesis-deficient DNAPs to cleave short hairpin structures.

Figure 19B shows an autoradiogram of a gel resolving the products of a cleavage reaction run using the substrate shown in Fig. 19A.

Figure 20A shows the A- and T-hairpin molecules used in the trigger/detection assay.

Figure 20B shows the sequence of the alpha primer used in the trigger/detection assay.

Figure 20C shows the structure of the cleaved A- and T-hairpin molecules.

Figure 20D depicts the complementarity between the A- and T-hairpin molecules.

Figure 21 provides the complete 206-mer duplex sequence employed as a substrate for the 5' nucleases of the present invention

Figures 22A and B show the cleavage of linear nucleic acid substrates (based on the 206-mer of Figure 21) by wild type DNAPs and 5' nucleases isolated from *Thermus aquaticus* and *Thermus flavus*.

Figure 23 provides a detailed schematic corresponding to the of one embodiment of the detection method of the present invention.

Figure 24 shows the propagation of cleavage of the linear duplex nucleic acid structures of Figure 23 by the 5' nucleases of the present invention.

Figure 25A shows the "nibbling" phenomenon detected with the DNAPs of the present invention.

Figure 25B shows that the "nibbling" of Figure 25A is 5' nucleolytic cleavage and not phosphatase cleavage.

Figure 26 demonstrates that the "nibbling" phenomenon is duplex dependent.

Figure 27 is a schematic showing how "nibbling" can be employed in a detection assay.

Figure 28 demonstrates that "nibbling" can be target directed.

Figure 29 is a schematic showing the CFLP™ method of generating a characteristic fingerprint from a nucleic acid substrate.

Figure 30 shows an autoradiograph of a gel resolving the products of cleavage reactions run in the presence of either $MgCl_2$ or $MnCl_2$.

Figure 31 shows an autoradiograph of a gel resolving the products of cleavage reactions run on four similarly sized DNA substrates.

Figure 32 shows an autoradiograph of a gel resolving the products of cleavage reactions run using a wild-type and two mutant tyrosinase gene substrates.

Figure 33 shows an autoradiograph of a gel resolving the products of cleavage reactions run using either a wild-type or mutant tyrosinase substrate varying in length from 157 nucleotides to 1.587 kb.

Figure 34 shows an autoradiograph of a gel resolving the products of cleavage reactions run in various concentrations of $MnCl_2$.

Figure 35 shows an autoradiograph of a gel resolving the products of cleavage reactions run in various concentrations of KCl.

Figure 36 shows an autoradiograph of a gel resolving the products of cleavage reactions run for different lengths of time.

Figure 37 shows an autoradiograph of a gel resolving the products of cleavage reactions run at different temperatures.

Figure 38 shows an autoradiograph of a gel resolving the products of cleavage reactions run using different amounts of the enzyme Cleavase™ BN.

Figure 39 shows an autoradiograph of a gel resolving the products of cleavage reactions run using four different preparations of the DNA substrate.

Figure 40 shows an autoradiograph of a gel resolving the products of cleavage reactions run on either the sense or antisense strand of four different tyrosinase gene substrates.

Figure 41 shows an autoradiograph of a gel resolving the products of cleavage reactions run on a wild-type β-globin substrate in two different concentrations of KCl and at four different temperatures.

Figure 42 shows an autoradiograph of a gel resolving the products of cleavage reactions run on two different mutant β-globin substrates in five different concentrations of KCl.

Figure 43 shows an autoradiograph of a gel resolving the products of cleavage reactions run on a wild-type and three mutant β-globin substrates.

Figure 44 shows an autoradiograph of a gel resolving the products of cleavage reactions run on an RNA substrate.

Figure 45 shows an autoradiograph of a gel resolving the products of cleavage reactions run using either the enzyme Cleavase™ BN or *Taq* DNA polymerase as the 5' nuclease.

Figure 46 shows an autoradiograph of a gel resolving the products of cleavage reactions run on a double-stranded DNA substrate to demonstrate multiplexing of the cleavage reaction.

Figure 47 shows an autoradiograph of a gel resolving the products of cleavage reactions run on double-stranded DNA substrates consisting of the 419 and 422 mutant alleles derived from exon 4 of the human tyrosinase gene in the presence of various concentrations of $MnCl_2$.

Figure 48 displays two traces representing two channel signals (JOE and FAM fluorescent dyes) for cleavage fragments derived from a cleavage reaction containing two differently labelled substrates (the wild-type and 422 mutant substrates derived from exon 4 of the tyrosinase gene). The thin lines represent the JOE-labelled wild-type substrate and the thick lines represent the FAM-labelled 422 mutant substrate. Above the tracing is an autoradiograph of a gel resolving the products of cleavage reactions run on double-stranded DNA substrates consisting of the wild-type and 422 mutant alleles derived from exon 4 of the tyrosinase gene.

Figure 49 depicts the nucleotide sequence of six SIV LTR clones corresponding to SEQ ID NOS:76-81.

Figure 50 shows an autoradiograph of a gel resolving the products of cleavage reactions run on six different double-stranded SIV LTR substrates which contained a biotin label on the 5' end of the (-) strand.

Figure 51 shows an autoradiograph of a gel resolving the products of cleavage reactions run on six different double-stranded SIV LTR substrates which contained a biotin label on the 5' end of the (+) strand.

Figure 52 shows an autoradiograph of a gel resolving the products of single-stranded cleavage reactions run in various concentrations of NaCl.

Figure 53 shows an autoradiograph of a gel resolving the products of single-stranded cleavage reactions run in various concentrations of $(NH_4)_2SO_4$.

Figure 54 shows an autoradiograph of a gel resolving the products of single-stranded cleavage reactions run in increasing concentrations of KCl.

Figure 55 shows an autoradiograph of a gel resolving the products of single-stranded cleavage reactions run in two concentrations of KCl for various periods of time.

Figure 56 shows an autoradiograph of a gel resolving the products of cleavage reactions run on either the single-stranded or double-stranded form of the same substrate.

Figure 57 shows an autoradiograph of a gel resolving the products of double-stranded cleavage reactions run in various concentrations of KCl.

Figure 58 shows an autoradiograph of a gel resolving the products of double-stranded cleavage reactions run in various concentrations of NaCl.

Figure 59 shows an autoradiograph of a gel resolving the products of double-stranded cleavage reactions run in various concentrations of $(NH_4)_2SO_4$.

Figure 60 shows an autoradiograph of a gel resolving the products of double-stranded cleavage reactions run for various lengths of time.

Figure 61 shows an autoradiograph of a gel resolving the products of double-stranded cleavage reactions run using various amounts of Cleavase™ BN enzyme for either 5 seconds or 1 minute.

Figure 62 shows an autoradiograph of a gel resolving the products of double-stranded cleavage reactions run at various temperatures.

Figure 63 shows an autoradiograph of a gel resolving the products of double-stranded cleavage reactions run using various amounts of Cleavase™ BN enzyme.

Figure 64A shows an autoradiograph of a gel resolving the products of single-stranded cleavage reactions run in buffers having various pHs.

Figure 64B shows an autoradiograph of a gel resolving the products of single-stranded cleavage reactions run in buffers having a pH of either 7.5 or 7.8.

Figure 65A shows an autoradiograph of a gel resolving the products of double-stranded cleavage reactions run in buffers having a pH of either 8.2 or 7.2.

Figure 65B shows an autoradiograph of a gel resolving the products of double-stranded cleavage reactions run in buffers having a pH of either 7.5 or 7.8.

Figure 66 shows an autoradiograph of a gel resolving the products of single-stranded cleavage reactions run in the presence of various amounts of human genomic DNA.

Figure 67 shows an autoradiograph of a gel resolving the products of single-stranded cleavage reactions run using the *Tfl* DNA polymerase in two different concentrations of KCl.

Figure 68 shows an autoradiograph of a gel resolving the products of single-stranded cleavage reactions run using the *Tth* DNA polymerase in two different concentrations of KCl.

Figure 69 shows an autoradiograph of a gel resolving the products of single-stranded cleavage reactions run using the *E. coli* Exo III enzyme in two different concentrations of KCl.

Figure 70 shows an autoradiograph of a gel resolving the products of single-stranded cleavage reactions run on three different tyrosinase gene substrates (SEQ ID NOS:47, 54 and 55) using either the *Tth* DNA polymerase, the *E. coli* Exo III enzyme or Cleavase™ BN.

Figure 71 is a schematic drawing depicting the location of the 5' and 3' cleavage sites on a cleavage structure.

Figure 72 shows an autoradiograph of a gel resolving the products of single-stranded cleavage reactions run on three different tyrosinase gene substrates (SEQ ID NOS:47, 54 and 55) using either Cleavase™ BN or the Rad1/Rad10 complex.

Figure 73 shows an autoradiograph of a gel resolving the products of double-stranded cleavage reactions run on a wild-type and two mutant β-globin substrates.

Figure 74A shows an autoradiograph of a gel resolving the products of single-stranded cleavage reactions run on a wild-type and three mutant β-globin substrates.

Figure 74B shows an autoradiograph of a gel resolving the products of single-stranded cleavage reactions run on five mutant β-globin substrates.

Figure 75 shows an autoradiograph of a gel resolving the products of double-stranded cleavage reactions which varied the order of addition of the reaction components.

Figure 76 depicts the organization of the human p53 gene; exons are represented by the solid black boxes and are labelled 1-11. Five hot spot regions are shown as a blow-up of the region spanning exons 5-8; the hot spot regions are labelled A, A', B, C, and D.

Figure 77 provides a schematic showing the use of a first 2-step PCR technique for the generation DNA fragments containing p53 mutations.

Figure 78 provides a schematic showing the use of a second 2-step PCR technique for the generation DNA fragments containing p53 mutations.

Figure 79 shows an autoradiograph of a gel resolving the products of cleavage reactions run on a wild-type and two mutant p53 substrates.

Figure 80 shows an autoradiograph of a gel resolving the products of cleavage reactions run on a wild-type and three mutant p53 substrates.

Figure 81 shows an autoradiograph of a gel resolving the products of cleavage reactions run on a wild-type and a mutant p53 substrate where the mutant and wild-type substrates are present in various concentrations relative to one another.

Figure 82 provides an alignment of HCV clones 1.1 (SEQ ID NO:121), HCV2.1 (SEQ ID NO:122), HCV3.1 (SEQ ID NO:123), HCV4.2 (SEQ ID NO:124), HCV6.1 (SEQ ID NO:125) and HCV7.1 (SEQ ID NO:126).

Figure 83 shows a fluoroimager scan of a gel resolving the products of cleavage reactions run on six double-stranded HCV substrates labeled on either the sense or anti-sense strand.

Figure 84 shows an autoradiogram of a gel resolving the products of cleavage reactions run on a wild-type and two mutant *M. tuberculosis rpoB* substrates.

Figure 85A shows a fluoroimager scan of a gel resolving the products of cleavage reactions run on a wild-type and two mutant *M. tuberculosis rpoB* substrates prepared using either dTTP or dUTP.

Figure 85B shows a fluoroimager scan of the gel shown in Figure 85A following a longer period of electrophoresis.

Figure 86 shows an autoradiogram of a gel resolving the products of cleavage reactions run on a wild-type and three mutant *M. tuberculosis katG* substrates labeled on the sense strand.

Figure 87 shows a fluoroimager scan of a gel resolving the products of cleavage reactions run on a wild-type and three mutant *M. tuberculosis katG* substrates labeled on the anti-sense strand.

Figure 88 shows the location of primers along the sequence of the *E. coli rrsE* gene (SEQ ID NO:158).

Figure 89 provides an alignment of the *E. coli rrsE* (SEQ ID NO:158), *Cam.jejuni5* (SEQ ID NO:159), and *Stp.aureus* (SEQ ID NO:160) rRNA genes with the location of consensus PCR rRNA primers indicated in bold type.

Figure 90 shows a fluoroimager scan of a gel resolving the products of cleavage reactions run on four bacterial 16S rRNA substrates.

Figure 91A shows a fluoroimager scan of a gel resolving the products of cleavage reactions run on five bacterial 16S rRNA substrates.

Figure 91B shows bacterial a fluoroimager scan of a gel resolving the products of cleavage reactions run on five bacterial 16S rRNA substrates.

Figure 92 shows bacterial a fluoroimager scan of a gel resolving the products of cleavage reactions run on various bacterial 16S rRNA substrates.

Figure 93 shows bacterial a fluoroimager scan of a gel resolving the products of cleavage reactions run on eight bacterial 16S rRNA substrates.

Figure 94 shows an autoradiogram of a gel resolving the products of cleavage reactions run on a wild-type and mutant tyrosinase gene substrates prepared using naturally occurring deoxynucleotides or deoxynucleotide analogs.

## DEFINITIONS

To facilitate understanding of the invention, a number of terms are defined below.

The term "gene" refers to a DNA sequence that comprises control and coding sequences necessary for the production of a polypeptide or precursor. The polypeptide can be encoded by a full length coding sequence or by any portion of the coding sequence so long as the desired enzymatic activity is retained.

The term "wild-type" refers to a gene or gene product which has the characteristics of that gene or gene product when isolated from a naturally occurring source. A wild-type gene is that which is most frequently observed in a population and is thus arbitrarily designed the "normal" or "wild-type" form of the gene. In contrast, the term "modified" or "mutant" refers to a gene or gene product which displays modifications in sequence and or functional properties (*i.e.*, altered characteristics) when compared to the wild-type gene or gene product. It is noted that naturally-occurring mutants can be isolated; these are identified by the fact that they have altered characteristics when compared to the wild-type gene or gene product.

The term "recombinant DNA vector" as used herein refers to DNA sequences containing a desired coding sequence and appropriate DNA sequences necessary for the expression of the operably linked coding sequence in a particular host organism. DNA sequences necessary for expression in procaryotes include a promoter, optionally an operator sequence, a ribosome binding site and possibly other sequences. Eukaryotic cells are known to utilize promoters, polyadenlyation signals and enhancers.

The term "LTR" as used herein refers to the long terminal repeat found at each end of a provirus (*i.e.*, the integrated form of a retrovirus). The LTR contains numerous regulatory signals including transcriptional control elements, polyadenylation signals and sequences needed for replication and integration of the viral genome. The viral LTR is divided into three regions called U3, R and U5.

The U3 region contains the enhancer and promoter elements. The U5 region contains the polyadenylation signals. The R (repeat) region separates the U3 and U5 regions and transcribed sequences of the R region appear at both the 5' and 3' ends of the viral RNA.

The term "oligonucleotide" as used herein is defined as a molecule comprised of two or more deoxyribonucleotides or ribonucleotides, preferably more than three, and usually more than ten. The exact size will depend on many factors, which in turn depends on the ultimate function or use of the oligonucleotide. The oligonucleotide may be generated in any manner, including chemical synthesis, DNA replication, reverse transcription, or a combination thereof.

Because mononucleotides are reacted to make oligonucleotides in a manner such that the 5' phosphate of one mononucleotide pentose ring is attached to the 3' oxygen of its neighbor in one direction via a phosphodiester linkage, an end of an oligonucleotide is referred to as the "5' end" if its 5' phosphate is not linked to the 3' oxygen of a mononucleotide pentose ring and as the "3' end" if its 3' oxygen is not linked to a 5' phosphate of a subsequent mononucleotide pentose ring. As used herein, a nucleic acid sequence, even if internal to a larger oligonucleotide, also may be said to have 5' and 3' ends.

When two different, non-overlapping oligonucleotides anneal to different regions of the same linear complementary nucleic acid sequence, and the 3' end of one oligonucleotide points towards the 5' end of the other, the former may be called the "upstream" oligonucleotide and the latter the "downstream" oligonucleotide.

The term "primer" refers to an oligonucleotide which is capable of acting as a point of initiation of synthesis when placed under conditions in which primer extension

is initiated. An oligonucleotide "primer" may occur naturally, as in a purified restriction digest or may be produced synthetically.

A primer is selected to be "substantially" complementary to a strand of specific sequence of the template. A primer must be sufficiently complementary to hybridize with a template strand for primer elongation to occur. A primer sequence need not reflect the exact sequence of the template. For example, a non-complementary nucleotide fragment may be attached to the 5' end of the primer, with the remainder of the primer sequence being substantially complementary to the strand. Non-complementary bases or longer sequences can be interspersed into the primer, provided that the primer sequence has sufficient complementarity with the sequence of the template to hybridize and thereby form a template primer complex for synthesis of the extension product of the primer.

"Hybridization" methods involve the annealing of a complementary sequence to the target nucleic acid (the sequence to be detected). The ability of two polymers of nucleic acid containing complementary sequences to find each other and anneal through base pairing interaction is a well-recognized phenomenon. The initial observations of the "hybridization" process by Marmur and Lane, *Proc. Natl. Acad. Sci. USA* 46:453 (1960) and Doty *et al.*, *Proc. Natl. Acad. Sci. USA* 46:461 (1960) have been followed by the refinement of this process into an essential tool of modern biology. Nonetheless, a number of problems have prevented the wide scale use of hybridization as a tool in human diagnostics. Among the more formidable problems are: 1) the inefficiency of hybridization; 2) the low concentration of specific target sequences in a mixture of genomic DNA; and 3) the hybridization of only partially complementary probes and targets.

With regard to efficiency, it is experimentally observed that only a fraction of the possible number of probe-target complexes are formed in a hybridization reaction. This is particularly true with short oligonucleotide probes (less than 100 bases in length). There are three fundamental causes: a) hybridization cannot occur because of secondary and tertiary structure interactions; b) strands of DNA containing the target sequence have rehybridized (reannealed) to their complementary strand; and c) some

target molecules are prevented from hybridization when they are used in hybridization formats that immobilize the target nucleic acids to a solid surface.

Even where the sequence of a probe is completely complementary to the sequence of the target, *i.e.*, the target's primary structure, the target sequence must be made accessible to the probe via rearrangements of higher-order structure. These higher-order structural rearrangements may concern either the secondary structure or tertiary structure of the molecule. Secondary structure is determined by intramolecular bonding. In the case of DNA or RNA targets this consists of hybridization within a single, continuous strand of bases (as opposed to hybridization between two different strands). Depending on the extent and position of intramolecular bonding, the probe can be displaced from the target sequence preventing hybridization.

Solution hybridization of oligonucleotide probes to denatured double-stranded DNA is further complicated by the fact that the longer complementary target strands can renature or reanneal. Again, hybridized probe is displaced by this process. This results in a low yield of hybridization (low "coverage") relative to the starting concentrations of probe and target.

With regard to low target sequence concentration, the DNA fragment containing the target sequence is usually in relatively low abundance in genomic DNA. This presents great technical difficulties; most conventional methods that use oligonucleotide probes lack the sensitivity necessary to detect hybridization at such low levels.

One attempt at a solution to the target sequence concentration problem is the amplification of the detection signal. Most often this entails placing one or more labels on an oligonucleotide probe. In the case of non-radioactive labels, even the highest affinity reagents have been found to be unsuitable for the detection of single copy genes in genomic DNA with oligonucleotide probes. *See* Wallace *et al.*, *Biochimie* 67:755 (1985). In the case of radioactive oligonucleotide probes, only extremely high specific activities are found to show satisfactory results. *See* Studencki and Wallace, *DNA* 3:1 (1984) and Studencki *et al.*, *Human Genetics* 37:42 (1985).

With regard to complementarity, it is important for some diagnostic applications to determine whether the hybridization represents complete or partial complementarity. For example, where it is desired to detect simply the presence or absence of pathogen DNA (such as from a virus, bacterium, fungi, mycoplasma, protozoan) it is only important that the hybridization method ensures hybridization when the relevant sequence is present; conditions can be selected where both partially complementary probes and completely complementary probes will hybridize. Other diagnostic applications, however, may require that the hybridization method distinguish between partial and complete complementarity. It may be of interest to detect genetic polymorphisms. For example, human hemoglobin is composed, in part, of four polypeptide chains. Two of these chains are identical chains of 141 amino acids (alpha chains) and two of these chains are identical chains of 146 amino acids (beta chains). The gene encoding the beta chain is known to exhibit polymorphism. The normal allele encodes a beta chain having glutamic acid at the sixth position. The mutant allele encodes a beta chain having valine at the sixth position. This difference in amino acids has a profound (most profound when the individual is homozygous for the mutant allele) physiological impact known clinically as sickle cell anemia. It is well known that the genetic basis of the amino acid change involves a single base difference between the normal allele DNA sequence and the mutant allele DNA sequence.

Unless combined with other techniques (such as restriction enzyme analysis), methods that allow for the same level of hybridization in the case of both partial as well as complete complementarity are typically unsuited for such applications; the probe will hybridize to both the normal and variant target sequence. Hybridization, regardless of the method used, requires some degree of complementarity between the sequence being assayed (the target sequence) and the fragment of DNA used to perform the test (the probe). (Of course, one can obtain binding without any complementarity but this binding is nonspecific and to be avoided.)

The complement of a nucleic acid sequence as used herein refers to an oligonucleotide which, when aligned with the nucleic acid sequence such that the 5'

end of one sequence is paired with the 3' end of the other, is in "antiparallel association." Certain bases not commonly found in natural nucleic acids may be included in the nucleic acids of the present invention and include, for example, inosine and 7-deazaguanine. Complementarity need not be perfect; stable duplexes may contain mismatched base pairs or unmatched bases. Those skilled in the art of nucleic acid technology can determine duplex stability empirically considering a number of variables including, for example, the length of the oligonucleotide, base composition and sequence of the oligonucleotide, ionic strength and incidence of mismatched base pairs.

Stability of a nucleic acid duplex is measured by the melting temperature, or "$T_m$." The $T_m$ of a particular nucleic acid duplex under specified conditions is the temperature at which on average half of the base pairs have disassociated.

The term "probe" as used herein refers to a labeled oligonucleotide which forms a duplex structure with a sequence in another nucleic acid, due to complementarity of at least one sequence in the probe with a sequence in the other nucleic acid.

The term "label" as used herein refers to any atom or molecule which can be used to provide a detectable (preferably quantifiable) signal, and which can be attached to a nucleic acid or protein. Labels may provide signals detectable by fluorescence, radioactivity, colorimetry, gravimetry, X-ray diffraction or absorption, magnetism, enzymatic activity, and the like.

The term "cleavage structure" as used herein, refers to a region of a single-stranded nucleic acid substrate containing secondary structure, said region being cleavable by a cleavage means, including but not limited to an enzyme. The cleavage structure is a substrate for specific cleavage by said cleavage means in contrast to a nucleic acid molecule which is a substrate for non-specific cleavage by agents such as phosphodiesterases which cleave nucleic acid molecules without regard to secondary structure (*i.e.*, no folding of the substrate is required).

The term "cleavage means" as used herein refers to any means which is capable of cleaving a cleavage structure, including but not limited to enzymes. The cleavage

means may include native DNAPs having 5' nuclease activity (*e.g.*, *Taq* DNA polymerase, *E. coli* DNA polymerase I) and, more specifically, modified DNAPs having 5' nuclease but lacking synthetic activity. The ability of 5' nucleases to cleave naturally occurring structures in nucleic acid templates (structure-specific cleavage) is useful to detect internal sequence differences in nucleic acids without prior knowledge of the specific sequence of the nucleic acid. In this manner, they are structure-specific enzymes. Structure-specific enzymes are enzymes which recognize specific secondary structures in a nucleic molecule and cleave these structures. The site of cleavage may be on either the 5' or 3' side of the cleavage structure; alternatively the site of cleavage may be between the 5' and 3' side (*i.e.*, within or internal to) of the cleavage structure. The cleavage means of the invention cleave a nucleic acid molecule in response to the formation of cleavage structures; it is not necessary that the cleavage means cleave the cleavage structure at any particular location within the cleavage structure.

The cleavage means is not restricted to enzymes having 5' nuclease activity. The cleavage means may include nuclease activity provided from a variety of sources including the enzyme Cleavase™, *Taq* DNA polymerase, *E. coli* DNA polymerase I and eukaryotic structure-specific endonucleases, murine FEN-1 endonucleases [Harrington and Liener, (1994) Genes and Develop. 8:1344] and calf thymus 5' to 3' exonuclease [Murante, R.S., *et al.* (1994) J. Biol. Chem. 269:1191]). In addition, enzymes having 3' nuclease activity such as members of the family of DNA repair endonucleases (*e.g.*, the *RrpI* enzyme from *Drosophila melanogaster*, the yeast RAD1/RAD10 complex and *E. coli* Exo III), are also suitable cleavage means for the practice of the methods of the invention.

The term "cleavage products" as used herein, refers to products generated by the reaction of a cleavage means with a cleavage structure (*i.e.*, the treatment of a cleavage structure with a cleavage means).

The terms "nucleic acid substrate" and nucleic acid template" are used herein interchangeably and refer to a nucleic acid molecule which when denatured and

allowed to renature (*i.e.*, to fold upon itself by the formation of intra-strand hydrogen bonds), forms at least one cleavage structure. The nucleic acid substrate may comprise single- or double-stranded DNA or RNA.

The term "substantially single-stranded" when used in reference to a nucleic acid substrate means that the substrate molecule exists primarily as a single strand of nucleic acid in contrast to a double-stranded substrate which exists as two strands of nucleic acid which are held together by inter-strand base pairing interactions.

Nucleic acids form secondary structures which depend on base-pairing for stability. When single strands of nucleic acids (single-stranded DNA, denatured double-stranded DNA or RNA) with different sequences, even closely related ones, are allowed to fold on themselves, they assume characteristic secondary structures. At "elevated temperatures" the duplex regions of the structures are brought to the brink of instability, so that the effects of small changes in sequence are maximized, and revealed as alterations in the cleavage pattern. In other words, "an elevated temperature" is a temperature at which a given duplex region of the folded substrate molecule is near the temperature at which that duplex melts. An alteration in the sequence of the substrate will then be likely to cause the destruction of a duplex region(s) thereby generating a different cleavage pattern when a cleavage agent which is dependent upon the recognition of structure is utilized in the reaction. While not being limited to any particular theory, it is thought that individual molecules in the target (*i.e.*, the substrate) population may each assume only one or a few of the potential cleavage structures (*i.e.*, duplexed regions), but when the sample is analyzed as a whole, a composite pattern representing all cleavage sites is detected. Many of the structures recognized as active cleavage sites are likely to be only a few base-pairs long and would appear to be unstable when elevated temperatures used in the cleavage reaction. Nevertheless, transient formation of these structures allows recognition and cleavage of these structures by said cleavage means. The formation or disruption of these structures in response to small sequence changes results in changes in the patterns of cleavage. Temperatures in the range of 40-85°C, with the range of 55-

85°C being particularly preferred, are suitable elevated temperatures for the practice of the method of the invention.

The term "sequence variation" as used herein refers to differences in nucleic acid sequence between two nucleic acid templates. For example, a wild-type structural gene and a mutant form of this wild-type structural gene may vary in sequence by the presence of single base substitutions and/or deletions or insertions of one or more nucleotides. These two forms of the structural gene are said to vary in sequence from one another. A second mutant form of the structural gene may exits. This second mutant form is said to vary in sequence from both the wild-type gene and the first mutant form of the gene. It is noted, however, that the invention does not require that a comparison be made between one or more forms of a gene to detect sequence variations. Because the method of the invention generates a characteristic and reproducible pattern of cleavage products for a given nucleic acid substrate, a characteristic "fingerprint" may be obtained from any nucleic substrate without reference to a wild-type or other control. The invention contemplates the use of the method for both "fingerprinting" nucleic acids without reference to a control and identification of mutant forms of a substrate nucleic acid by comparison of the mutant form of the substrate with a wild-type or known mutant control.

The term "liberating" as used herein refers to the release of a nucleic acid fragment from a larger nucleic acid fragment, such as an oligonucleotide, by the action of a 5' nuclease such that the released fragment is no longer covalently attached to the remainder of the oligonucleotide.

The term "substrate strand" as used herein, means that strand of nucleic acid in a cleavage structure in which the cleavage mediated by the 5' nuclease activity occurs.

The term "template strand" as used herein, means that strand of nucleic acid in a cleavage structure which is at least partially complementary to the substrate strand and which anneals to the substrate strand to form the cleavage structure.

The term "$K_m$" as used herein refers to the Michaelis-Menten constant for an enzyme and is defined as the concentration of the specific substrate at which a given enzyme yields one-half its maximum velocity in an enzyme catalyzed reaction.

The term "nucleotide analog" as used herein refers to modified or non-naturally occurring nucleotides such as 7-deaza purines (*i.e.*, 7-deaza-dATP and 7-deaza-dGTP). Nucleotide analogs include base analogs and comprise modified forms of deoxyribonucleotides as well as ribonucleotides. As used herein the term "nucleotide analog" when used in reference to substrates present in a PCR mixture refers to the use of nucleotides other than dATP, dGTP, dCTP and dTTP; thus, the use of dUTP (a naturally occurring dNTP) in a PCR would comprise the use of a nucleotide analog in the PCR. A PCR product generated using dUTP, 7-deaza-dATP, 7-deaza-dGTP or any other nucleotide analog in the reaction mixture is said to contain nucleotide analogs.

"Oligonucleotide primers matching or complementary to a gene sequence" refers to oligonucleotide primers capable of facilitating the template-dependent synthesis of single or double-stranded nucleic acids. Oligonucleotide primers matching or complementary to a gene sequence may be used in PCRs, RT-PCRs and the like.

A "consensus gene sequence" refers to a gene sequence which is derived by comparison of two or more gene sequences and which describes the nucleotides most often present in a given segment of the genes; the consensus sequence is the canonical sequence.

The term "polymorphic locus" is a locus present in a population which shows variation between members of the population (*i.e.*, the most common allele has a frequency of less than 0.95). In contrast, a "monomorphic locus" is a genetic locus at little or no variations seen between members of the population (generally taken to be a locus at which the most common allele exceeds a frequency of 0.95 in the gene pool of the population).

The term "microorganism" as used herein means an organism too small to be observed with the unaided eye and includes, but is not limited to bacteria, virus, protozoans, fungi, and ciliates.

The term "microbial gene sequences" refers to gene sequences derived from a microorganism.

The term "bacteria" refers to any bacterial species including eubacterial and archaebacterial species.

The term "virus" refers to obligate, ultramicroscopic, intracellular parasites incapable of autonomous replication (*i.e.*, replication requires the use of the host cell's machinery).

The term "multi-drug resistant" or multiple-drug resistant" refers to a microorganism which is resistant to more than one of the antibiotics or antimicrobial agents used in the treatment of said microorganism.

## DESCRIPTION OF THE INVENTION

The present invention relates to methods and compositions for treating nucleic acid, and in particular, methods and compositions for detection and characterization of nucleic acid sequences and sequence changes.

The present invention relates to means for cleaving a nucleic acid cleavage structure in a site-specific manner. In particular, the present invention relates to a cleaving enzyme having 5' nuclease activity without interfering nucleic acid synthetic ability.

This invention provides 5' nucleases derived from thermostable DNA polymerases which exhibit altered DNA synthetic activity from that of native thermostable DNA polymerases. The 5' nuclease activity of the polymerase is retained while the synthetic activity is reduced or absent. Such 5' nucleases are capable of catalyzing the structure-specific cleavage of nucleic acids in the absence of interfering synthetic activity. The lack of synthetic activity during a cleavage reaction results in nucleic acid cleavage products of uniform size.

The novel properties of the polymerases of the invention form the basis of a method of detecting specific nucleic acid sequences. This method relies upon the amplification of the detection molecule rather than upon the amplification of the target sequence itself as do existing methods of detecting specific target sequences.

DNA polymerases (DNAPs), such as those isolated from *E. coli* or from thermophilic bacteria of the genus *Thermus*, are enzymes that synthesize new DNA strands. Several of the known DNAPs contain associated nuclease activities in addition to the synthetic activity of the enzyme.

Some DNAPs are known to remove nucleotides from the 5' and 3' ends of DNA chains [Kornberg, *DNA Replication*, W.H. Freeman and Co., San Francisco, pp. 127-139 (1980)]. These nuclease activities are usually referred to as 5' exonuclease and 3' exonuclease activities, respectively. For example, the 5' exonuclease activity located in the N-terminal domain of several DNAPs participates in the removal of RNA primers during lagging strand synthesis during DNA replication and the removal of damaged nucleotides during repair. Some DNAPs, such as the *E. coli* DNA polymerase (DNAPEc1), also have a 3' exonuclease activity responsible for proof-reading during DNA synthesis (Kornberg, *supra*).

A DNAP isolated from *Thermus aquaticus*, termed *Taq* DNA polymerase (DNAP*Taq*), has a 5' exonuclease activity, but lacks a functional 3' exonucleolytic domain [Tindall and Kunkell, *Biochem.* 27:6008 (1988)]. Derivatives of DNAPEc1 and DNAP*Taq*, respectively called the Klenow and Stoffel fragments, lack 5' exonuclease domains as a result of enzymatic or genetic manipulations [Brutlag *et al.*, *Biochem. Biophys. Res. Commun.* 37:982 (1969); Erlich *et al.*, *Science* 252:1643 (1991); Setlow and Kornberg, *J. Biol. Chem.* 247:232 (1972)].

The 5' exonuclease activity of DNAP*Taq* was reported to require concurrent synthesis [Gelfand, *PCR Technology - Principles and Applications for DNA Amplification* (H.A. Erlich, Ed.), Stockton Press, New York, p. 19 (1989)]. Although mononucleotides predominate among the digestion products of the 5' exonucleases of DNAP*Taq* and DNAPEc1, short oligonucleotides (≤ 12 nucleotides) can also be observed implying that these so-called 5' exonucleases can function endonucleolytically [Setlow, *supra*; Holland *et al.*, *Proc. Natl. Acad. Sci.* USA 88:7276 (1991)].

In WO 92/06200, Gelfand *et al.* show that the preferred substrate of the 5' exonuclease activity of the thermostable DNA polymerases is displaced single-stranded DNA. Hydrolysis of the phosphodiester bond occurs between the displaced single-stranded DNA and the double-helical DNA with the preferred exonuclease cleavage site being a phosphodiester bond in the double helical region. Thus, the 5'

exonuclease activity usually associated with DNAPs is a structure-dependent single-stranded endonuclease and is more properly referred to as a 5' nuclease. Exonucleases are enzymes which cleave nucleotide molecules from the ends of the nucleic acid molecule. Endonucleases, on the other hand, are enzymes which cleave the nucleic acid molecule at internal rather than terminal sites. The nuclease activity associated with some thermostable DNA polymerases cleaves endonucleolytically but this cleavage requires contact with the 5' end of the molecule being cleaved. Therefore, these nucleases are referred to as 5' nucleases.

When a 5' nuclease activity is associated with a eubacterial Type A DNA polymerase, it is found in the one-third N-terminal region of the protein as an independent functional domain. The C-terminal two-thirds of the molecule constitute the polymerization domain which is responsible for the synthesis of DNA. Some Type A DNA polymerases also have a 3' exonuclease activity associated with the two-third C-terminal region of the molecule.

The 5' exonuclease activity and the polymerization activity of DNAPs have been separated by proteolytic cleavage or genetic manipulation of the polymerase molecule. To date thermostable DNAPs have been modified to remove or reduce the amount of 5' nuclease activity while leaving the polymerase activity intact.

The Klenow or large proteolytic cleavage fragment of DNAPEc1 contains the polymerase and 3' exonuclease activity but lacks the 5' nuclease activity. The Stoffel fragment of DNAP*Taq* (DNAPStf) lacks the 5' nuclease activity due to a genetic manipulation which deleted the N-terminal 289 amino acids of the polymerase molecule [Erlich *et al.*, *Science* 252:1643 (1991)]. WO 92/06200 describes a thermostable DNAP with an altered level of 5' to 3' exonuclease. U.S. Patent No. 5,108,892 describes a *Thermus aquaticus* DNAP without a 5' to 3' exonuclease. However, the art of molecular biology lacks a thermostable DNA polymerase with a lessened amount of synthetic activity.

The present invention provides 5' nucleases derived from thermostable Type A DNA polymerases that retain 5' nuclease activity but have reduced or absent synthetic activity. The ability to uncouple the synthetic activity of the enzyme from the 5'

nuclease activity proves that the 5' nuclease activity does not require concurrent DNA synthesis as was previously reported (Gelfand, *PCR Technology, supra*).

The description of the invention is divided into: I. Detection of Specific Nucleic Acid Sequences Using 5' Nucleases; II. Generation of 5' Nucleases Derived From Thermostable DNA Polymerases; III. Therapeutic Uses of 5' Nucleases; IV. Detection of Antigenic or Nucleic Acid Targets by a Dual Capture Assay; and V. Cleavase™ Fragment Length Polymorphism for the Detection of Secondary Structure and VI. Detection of Mutations in the p53 Tumor Suppressor Gene Using the CFLP™ Method.

## I. Detection Of Specific Nucleic Acid Sequences Using 5' Nucleases

The 5' nucleases of the invention form the basis of a novel detection assay for the identification of specific nucleic acid sequences. This detection system identifies the presence of specific nucleic acid sequences by requiring the annealing of two oligonucleotide probes to two portions of the target sequence. As used herein, the term "target sequence" or "target nucleic acid sequence" refers to a specific nucleic acid sequence within a polynucleotide sequence, such as genomic DNA or RNA, which is to be either detected or cleaved or both.

Figure 1A provides a schematic of one embodiment of the detection method of the present invention. The target sequence is recognized by two distinct oligonucleotides in the triggering or trigger reaction. It is preferred that one of these oligonucleotides is provided on a solid support. The other can be provided free. In Figure 1A the free oligo is indicated as a "primer" and the other oligo is shown attached to a bead designated as type 1. The target nucleic acid aligns the two oligonucleotides for specific cleavage of the 5' arm (of the oligo on bead 1) by the DNAPs of the present invention (not shown in Figure 1A).

The site of cleavage (indicated by a large solid arrowhead) is controlled by the distance between the 3' end of the "primer" and the downstream fork of the oligo on bead 1. The latter is designed with an uncleavable region (indicated by the striping).

In this manner neither oligonucleotide is subject to cleavage when misaligned or when unattached to target nucleic acid.

Successful cleavage releases a single copy of what is referred to as the alpha signal oligo. This oligo may contain a detectable moiety (*e.g.*, fluorescein). On the other hand, it may be unlabelled.

In one embodiment of the detection method, two more oligonucleotides are provided on solid supports. The oligonucleotide shown in Figure 1A on bead 2 has a region that is complementary to the alpha signal oligo (indicated as alpha prime) allowing for hybridization. This structure can be cleaved by the DNAPs of the present invention to release the beta signal oligo. The beta signal oligo can then hybridize to type 3 beads having an oligo with a complementary region (indicated as beta prime). Again, this structure can be cleaved by the DNAPs of the present invention to release a new alpha oligo.

At this point, the amplification has been linear. To increase the power of the method, it is desired that the alpha signal oligo hybridized to bead type 2 be liberated after release of the beta oligo so that it may go on to hybridize with other oligos on type 2 beads. Similarly, after release of an alpha oligo from type 3 beads, it is desired that the beta oligo be liberated.

The liberation of "captured" signal oligos can be achieved in a number of ways. First, it has been found that the DNAPs of the present invention have a true 5' exonuclease capable of "nibbling" the 5' end of the alpha (and beta) prime oligo (discussed below in more detail). Thus, under appropriate conditions, the hybridization is destabilized by nibbling of the DNAP. Second, the alpha - alpha prime (as well as the beta - beta prime) complex can be destabilized by heat (*e.g.*, thermal cycling).

With the liberation of signal oligos by such techniques, each cleavage results in a doubling of the number of signal oligos. In this manner, detectable signal can quickly be achieved.

Figure 1B provides a schematic of a second embodiment of the detection method of the present invention. Again, the target sequence is recognized by two

distinct oligonucleotides in the triggering or trigger reaction and the target nucleic acid aligns the two oligonucleotides for specific cleavage of the 5' arm by the DNAPs of the present invention (not shown in Figure 1B). The first oligo is completely complementary to a portion of the target sequence. The second oligonucleotide is partially complementary to the target sequence; the 3' end of the second oligonucleotide is fully complementary to the target sequence while the 5' end is non-complementary and forms a single-stranded arm. The non-complementary end of the second oligonucleotide may be a generic sequence which can be used with a set of standard hairpin structures (described below). The detection of different target sequences would require unique portions of two oligonucleotides: the entire first oligonucleotide and the 3' end of the second oligonucleotide. The 5' arm of the second oligonucleotide can be invariant or generic in sequence.

The annealing of the first and second oligonucleotides near one another along the target sequence forms a forked cleavage structure which is a substrate for the 5' nuclease of DNA polymerases. The approximate location of the cleavage site is again indicated by the large solid arrowhead in Figure 1B.

The 5' nucleases of the invention are capable of cleaving this structure but are not capable of polymerizing the extension of the 3' end of the first oligonucleotide. The lack of polymerization activity is advantageous as extension of the first oligonucleotide results in displacement of the annealed region of the second oligonucleotide and results in moving the site of cleavage along the second oligonucleotide. If polymerization is allowed to occur to any significant amount, multiple lengths of cleavage product will be generated. A single cleavage product of uniform length is desirable as this cleavage product initiates the detection reaction.

The trigger reaction may be run under conditions that allow for thermocycling. Thermocycling of the reaction allows for a logarithmic increase in the amount of the trigger oligonucleotide released in the reaction.

The second part of the detection method allows the annealing of the fragment of the second oligonucleotide liberated by the cleavage of the first cleavage structure formed in the triggering reaction (called the third or trigger oligonucleotide) to a first

hairpin structure. This first hairpin structure has a single-stranded 5' arm and a single-stranded 3' arm. The third oligonucleotide triggers the cleavage of this first hairpin structure by annealing to the 3' arm of the hairpin thereby forming a substrate for cleavage by the 5' nuclease of the present invention. The cleavage of this first hairpin structure generates two reaction products: 1) the cleaved 5' arm of the hairpin called the fourth oligonucleotide, and 2) the cleaved hairpin structure which now lacks the 5' arm and is smaller in size than the uncleaved hairpin. This cleaved first hairpin may be used as a detection molecule to indicate that cleavage directed by the trigger or third oligonucleotide occurred. Thus, this indicates that the first two oligonucleotides found and annealed to the target sequence thereby indicating the presence of the target sequence in the sample.

The detection products are amplified by having the fourth oligonucleotide anneal to a second hairpin structure. This hairpin structure has a 5' single-stranded arm and a 3' single-stranded arm. The fourth oligonucleotide generated by cleavage of the first hairpin structure anneals to the 3' arm of the second hairpin structure thereby creating a third cleavage structure recognized by the 5' nuclease. The cleavage of this second hairpin structure also generates two reaction products: 1) the cleaved 5' arm of the hairpin called the fifth oligonucleotide which is similar or identical in sequence to the third nucleotide, and 2) the cleaved second hairpin structure which now lacks the 5' arm and is smaller in size than the uncleaved hairpin. This cleaved second hairpin may be as a detection molecule and amplifies the signal generated by the cleavage of the first hairpin structure. Simultaneously with the annealing of the forth oligonucleotide, the third oligonucleotide is dissociated from the cleaved first hairpin molecule so that it is free to anneal to a new copy of the first hairpin structure. The disassociation of the oligonucleotides from the hairpin structures may be accomplished by heating or other means suitable to disrupt base-pairing interactions.

Further amplification of the detection signal is achieved by annealing the fifth oligonucleotide (similar or identical in sequence to the third oligonucleotide) to another molecule of the first hairpin structure. Cleavage is then performed and the oligonucleotide that is liberated then is annealed to another molecule of the second

hairpin structure. Successive rounds of annealing and cleavage of the first and second hairpin structures, provided in excess, are performed to generate a sufficient amount of cleaved hairpin products to be detected. The temperature of the detection reaction is cycled just below and just above the annealing temperature for the oligonucleotides used to direct cleavage of the hairpin structures, generally about 55°C to 70°C. The number of cleavages will double in each cycle until the amount of hairpin structures remaining is below the $K_m$ for the hairpin structures. This point is reached when the hairpin structures are substantially used up. When the detection reaction is to be used in a quantitative manner, the cycling reactions are stopped before the accumulation of the cleaved hairpin detection products reach a plateau.

Detection of the cleaved hairpin structures may be achieved in several ways. In one embodiment detection is achieved by separation on agarose or polyacrylamide gels followed by staining with ethidium bromide. In another embodiment, detection is achieved by separation of the cleaved and uncleaved hairpin structures on a gel followed by autoradiography when the hairpin structures are first labelled with a radioactive probe and separation on chromatography columns using HPLC or FPLC followed by detection of the differently sized fragments by absorption at $OD_{260}$. Other means of detection include detection of changes in fluorescence polarization when the single-stranded 5' arm is released by cleavage, the increase in fluorescence of an intercalating fluorescent indicator as the amount of primers annealed to 3' arms of the hairpin structures increases. The formation of increasing amounts of duplex DNA (between the primer and the 3' arm of the hairpin) occurs if successive rounds of cleavage occur.

The hairpin structures may be attached to a solid support, such as an agarose, styrene or magnetic bead, via the 3' end of the hairpin. A spacer molecule may be placed between the 3' end of the hairpin and the bead, if so desired. The advantage of attaching the hairpin structures to a solid support is that this prevents the hybridization of the two hairpin structures to one another over regions which are complementary. If the hairpin structures anneal to one another, this would reduce the amount of hairpins available for hybridization to the primers released during the cleavage reactions. If the

hairpin structures are attached to a solid support, then additional methods of detection of the products of the cleavage reaction may be employed. These methods include, but are not limited to, the measurement of the released single-stranded 5' arm when the 5' arm contains a label at the 5' terminus. This label may be radioactive, fluorescent, biotinylated, etc. If the hairpin structure is not cleaved, the 5' label will remain attached to the solid support. If cleavage occurs, the 5' label will be released from the solid support.

The 3' end of the hairpin molecule may be blocked through the use of dideoxynucleotides. A 3' terminus containing a dideoxynucleotide is unavailable to participate in reactions with certain DNA modifying enzymes, such as terminal transferase. Cleavage of the hairpin having a 3' terminal dideoxynucleotide generates a new, unblocked 3' terminus at the site of cleavage. This new 3' end has a free hydroxyl group which can interact with terminal transferase thus providing another means of detecting the cleavage products.

The hairpin structures are designed so that their self-complementary regions are very short (generally in the range of 3-8 base pairs). Thus, the hairpin structures are not stable at the high temperatures at which this reaction is performed (generally in the range of 50-75°C) unless the hairpin is stabilized by the presence of the annealed oligonucleotide on the 3' arm of the hairpin. This instability prevents the polymerase from cleaving the hairpin structure in the absence of an associated primer thereby preventing false positive results due to non-oligonucleotide directed cleavage.

As discussed above, the use of the 5' nucleases of the invention which have reduced polymerization activity is advantageous in this method of detecting specific nucleic acid sequences. Significant amounts of polymerization during the cleavage reaction would cause shifting of the site of cleavage in unpredictable ways resulting in the production of a series of cleaved hairpin structures of various sizes rather than a single easily quantifiable product. Additionally, the primers used in one round of cleavage could, if elongated, become unusable for the next cycle, by either forming an incorrect structure or by being too long to melt off under moderate temperature cycling conditions. In a pristine system (*i.e.*, lacking the presence of dNTPs), one could use

the unmodified polymerase, but the presence of nucleotides (dNTPs) can decrease the per cycle efficiency enough to give a false negative result. When a crude extract (genomic DNA preparations, crude cell lysates, *etc.*) is employed or where a sample of DNA from a PCR reaction, or any other sample that might be contaminated with dNTPs, the 5' nucleases of the present invention that were derived from thermostable polymerases are particularly useful.

## II. Generation Of 5' Nucleases From Thermostable DNA Polymerases

The genes encoding Type A DNA polymerases share about 85% homology to each other on the DNA sequence level. Preferred examples of thermostable polymerases include those isolated from *Thermus aquaticus*, *Thermus flavus*, and *Thermus thermophilus*. However, other thermostable Type A polymerases which have 5' nuclease activity are also suitable. Figs. 2 and 3 compare the nucleotide and amino acid sequences of the three above mentioned polymerases. In Figures 2 and 3, the consensus or majority sequence derived from a comparison of the nucleotide (Fig. 2) or amino acid (Fig. 3) sequence of the three thermostable DNA polymerases is shown on the top line. A dot appears in the sequences of each of these three polymerases whenever an amino acid residue in a given sequence is identical to that contained in the consensus amino acid sequence. Dashes are used to introduce gaps in order to maximize alignment between the displayed sequences. When no consensus nucleotide or amino acid is present at a given position, an "X" is placed in the consensus sequence. SEQ ID NOS:1-3 display the nucleotide sequences and SEQ ID NOS:4-6 display the amino acid sequences of the three wild-type polymerases. SEQ ID NO:1 corresponds to the nucleic acid sequence of the wild type *Thermus aquaticus* DNA polymerase gene isolated from the YT-1 strain [Lawyer *et al.*, *J. Biol. Chem.* 264:6427 (1989)]. SEQ ID NO:2 corresponds to the nucleic acid sequence of the wild type *Thermus flavus* DNA polymerase gene [Akhmetzjanov and Vakhitov, *Nucl. Acids Res.* 20:5839 (1992)]. SEQ ID NO:3 corresponds to the nucleic acid sequence of the wild type *Thermus thermophilus* DNA polymerase gene [Gelfand *et al.*, WO 91/09950

(1991)]. SEQ ID NOS:7-8 depict the consensus nucleotide and amino acid sequences, respectively for the above three DNAPs (also shown on the top row in Figs. 2 and 3).

The 5' nucleases of the invention derived from thermostable polymerases have reduced synthetic ability, but retain substantially the same 5' exonuclease activity as the native DNA polymerase. The term "substantially the same 5' nuclease activity" as used herein means that the 5' nuclease activity of the modified enzyme retains the ability to function as a structure-dependent single-stranded endonuclease but not necessarily at the same rate of cleavage as compared to the unmodified enzyme. Type A DNA polymerases may also be modified so as to produce an enzyme which has increases 5' nuclease activity while having a reduced level of synthetic activity. Modified enzymes having reduced synthetic activity and increased 5' nuclease activity are also envisioned by the present invention.

By the term "reduced synthetic activity" as used herein it is meant that the modified enzyme has less than the level of synthetic activity found in the unmodified or "native" enzyme. The modified enzyme may have no synthetic activity remaining or may have that level of synthetic activity that will not interfere with the use of the modified enzyme in the detection assay described below. The 5' nucleases of the present invention are advantageous in situations where the cleavage activity of the polymerase is desired, but the synthetic ability is not (such as in the detection assay of the invention).

As noted above, it is not intended that the invention be limited by the nature of the alteration necessary to render the polymerase synthesis deficient. The present invention contemplates a variety of methods, including but not limited to: 1) proteolysis; 2) recombinant constructs (including mutants); and 3) physical and/or chemical modification and/or inhibition.

### 1. Proteolysis

Thermostable DNA polymerases having a reduced level of synthetic activity are produced by physically cleaving the unmodified enzyme with proteolytic enzymes to produce fragments of the enzyme that are deficient in synthetic activity but retain 5'

nuclease activity. Following proteolytic digestion, the resulting fragments are separated by standard chromatographic techniques and assayed for the ability to synthesize DNA and to act as a 5' nuclease. The assays to determine synthetic activity and 5' nuclease activity are described below.

**2. Recombinant Constructs**

The examples below describe a preferred method for creating a construct encoding a 5' nuclease derived from a thermostable DNA polymerase. As the Type A DNA polymerases are similar in DNA sequence, the cloning strategies employed for the *Thermus aquaticus* and *flavus* polymerases are applicable to other thermostable Type A polymerases. In general, a thermostable DNA polymerase is cloned by isolating genomic DNA using molecular biological methods from a bacteria containing a thermostable Type A DNA polymerase. This genomic DNA is exposed to primers which are capable of amplifying the polymerase gene by PCR.

This amplified polymerase sequence is then subjected to standard deletion processes to delete the polymerase portion of the gene. Suitable deletion processes are described below in the examples.

The example below discusses the strategy used to determine which portions of the DNAP*Taq* polymerase domain could be removed without eliminating the 5' nuclease activity. Deletion of amino acids from the protein can be done either by deletion of the encoding genetic material, or by introduction of a translational stop codon by mutation or frame shift. In addition, proteolytic treatment of the protein molecule can be performed to remove segments of the protein.

In the examples below, specific alterations of the *Taq* gene were: a deletion between nucleotides 1601 and 2502 (the end of the coding region), a 4 nucleotide insertion at position 2043, and deletions between nucleotides 1614 and 1848 and between nucleotides 875 and 1778 (numbering is as in SEQ ID NO:1). These modified sequences are described below in the examples and at SEQ ID NOS:9-12.

Those skilled in the art understand that single base pair changes can be innocuous in terms of enzyme structure and function. Similarly, small additions and

deletions can be present without substantially changing the exonuclease or polymerase function of these enzymes.

Other deletions are also suitable to create the 5' nucleases of the present invention. It is preferable that the deletion decrease the polymerase activity of the 5' nucleases to a level at which synthetic activity will not interfere with the use of the 5' nuclease in the detection assay of the invention. Most preferably, the synthetic ability is absent. Modified polymerases are tested for the presence of synthetic and 5' nuclease activity as in assays described below. Thoughtful consideration of these assays allows for the screening of candidate enzymes whose structure is heretofore as yet unknown. In other words, construct "X" can be evaluated according to the protocol described below to determine whether it is a member of the genus of 5' nucleases of the present invention as defined functionally, rather than structurally.

In the example below, the PCR product of the amplified *Thermus aquaticus* genomic DNA did not have the identical nucleotide structure of the native genomic DNA and did not have the same synthetic ability of the original clone. Base pair changes which result due to the infidelity of DNAP*Taq* during PCR amplification of a polymerase gene are also a method by which the synthetic ability of a polymerase gene may be inactivated. The examples below and Figs. 4A and 5A indicate regions in the native *Thermus aquaticus* and *flavus* DNA polymerases likely to be important for synthetic ability. There are other base pair changes and substitutions that will likely also inactivate the polymerase.

It is not necessary, however, that one start out the process of producing a 5' nuclease from a DNA polymerase with such a mutated amplified product. This is the method by which the examples below were performed to generate the synthesis-deficient DNAP*Taq* mutants, but it is understood by those skilled in the art that a wild-type DNA polymerase sequence may be used as the starting material for the introduction of deletions, insertion and substitutions to produce a 5' nuclease. For example, to generate the synthesis-deficient DNAPTfl mutant, the primers listed in SEQ ID NOS:13-14 were used to amplify the wild type DNA polymerase gene from

*Thermus flavus* strain AT-62. The amplified polymerase gene was then subjected to restriction enzyme digestion to delete a large portion of the domain encoding the synthetic activity.

The present invention contemplates that the nucleic acid construct of the present invention be capable of expression in a suitable host. Those in the art know methods for attaching various promoters and 3' sequences to a gene structure to achieve efficient expression. The examples below disclose two suitable vectors and six suitable vector constructs. Of course, there are other promoter/vector combinations that would be suitable. It is not necessary that a host organism be used for the expression of the nucleic acid constructs of the invention. For example, expression of the protein encoded by a nucleic acid construct may be achieved through the use of a cell-free in vitro transcription/translation system. An example of such a cell-free system is the commercially available TnT™ Coupled Reticulocyte Lysate System (Promega Corporation, Madison, WI).

Once a suitable nucleic acid construct has been made, the 5' nuclease may be produced from the construct. The examples below and standard molecular biological teachings enable one to manipulate the construct by different suitable methods.

Once the 5' nuclease has been expressed, the polymerase is tested for both synthetic and nuclease activity as described below.

### 3. Physical And/Or Chemical Modification And/Or Inhibition

The synthetic activity of a thermostable DNA polymerase may be reduced by chemical and/or physical means. In one embodiment, the cleavage reaction catalyzed by the 5' nuclease activity of the polymerase is run under conditions which preferentially inhibit the synthetic activity of the polymerase. The level of synthetic activity need only be reduced to that level of activity which does not interfere with cleavage reactions requiring no significant synthetic activity.

As shown in the examples below, concentrations of $Mg^{++}$ greater than 5 mM inhibit the polymerization activity of the native DNAP*Taq*. The ability of the 5'

nuclease to function under conditions where synthetic activity is inhibited is tested by running the assays for synthetic and 5' nuclease activity, described below, in the presence of a range of $Mg^{++}$ concentrations (5 to 10 mM). The effect of a given concentration of $Mg^{++}$ is determined by quantitation of the amount of synthesis and cleavage in the test reaction as compared to the standard reaction for each assay.

The inhibitory effect of other ions, polyamines, denaturants, such as urea, formamide, dimethylsulfoxide, glycerol and non-ionic detergents (Triton X-100 and Tween-20), nucleic acid binding chemicals such as, actinomycin D, ethidium bromide and psoralens, are tested by their addition to the standard reaction buffers for the synthesis and 5' nuclease assays. Those compounds having a preferential inhibitory effect on the synthetic activity of a thermostable polymerase are then used to create reaction conditions under which 5' nuclease activity (cleavage) is retained while synthetic activity is reduced or eliminated.

Physical means may be used to preferentially inhibit the synthetic activity of a polymerase. For example, the synthetic activity of thermostable polymerases is destroyed by exposure of the polymerase to extreme heat (typically 96 to 100°C) for extended periods of time (greater than or equal to 20 minutes). While these are minor differences with respect to the specific heat tolerance for each of the enzymes, these are readily determined. Polymerases are treated with heat for various periods of time and the effect of the heat treatment upon the synthetic and 5' nuclease activities is determined.

## III. Therapeutic Utility Of 5' Nucleases

The 5' nucleases of the invention have not only the diagnostic utility discussed above, but additionally have therapeutic utility for the cleavage and inactivation of specific mRNAs inside infected cells. The mRNAs of pathogenic agents, such as viruses, bacteria, are targeted for cleavage by a synthesis-deficient DNA polymerase by the introduction of a oligonucleotide complementary to a given mRNA produced by the pathogenic agent into the infected cell along with the synthesis-deficient polymerase. Any pathogenic agent may be targeted by this method provided the

nucleotide sequence information is available so that an appropriate oligonucleotide may be synthesized. The synthetic oligonucleotide anneals to the complementary mRNA thereby forming a cleavage structure recognized by the modified enzyme. The ability of the 5' nuclease activity of thermostable DNA polymerases to cleave RNA-DNA hybrids is shown herein in Example 1D.

Liposomes provide a convenient delivery system. The synthetic oligonucleotide may be conjugated or bound to the nuclease to allow for co-delivery of these molecules. Additional delivery systems may be employed.

Inactivation of pathogenic mRNAs has been described using antisense gene regulation and using ribozymes (Rossi, U.S. Patent No. 5,144,019, hereby incorporated by reference). Both of these methodologies have limitations.

The use of antisense RNA to impair gene expression requires stoichiometric and therefore, large molar excesses of anti-sense RNA relative to the pathogenic RNA to be effective. Ribozyme therapy, on the other hand, is catalytic and therefore lacks the problem of the need for a large molar excess of the therapeutic compound found with antisense methods. However, ribozyme cleavage of a given RNA requires the presence of highly conserved sequences to form the catalytically active cleavage structure. This requires that the target pathogenic mRNA contain the conserved sequences (GAAAC $(X)_n$ GU) thereby limiting the number of pathogenic mRNAs that can be cleaved by this method. In contrast, the catalytic cleavage of RNA by the use of a DNA oligonucleotide and a 5' nuclease is dependent upon structure only; thus, virtually any pathogenic RNA sequence can be used to design an appropriate cleavage structure.

## IV. Detection Of Antigenic Or Nucleic Acid Targets By A Dual Capture Assay

The ability to generate 5' nucleases from thermostable DNA polymerases provides the basis for a novel means of detecting the presence of antigenic or nucleic acid targets. In this dual capture assay, the polymerase domains encoding the synthetic activity and the nuclease activity are covalently attached to two separate and distinct

antibodies or oligonucleotides. When both the synthetic and the nuclease domains are present in the same reaction and dATP, dTTP and a small amount of poly d(A-T) are provided, an enormous amount of poly d(A-T) is produced. The large amounts of poly d(A-T) are produced as a result of the ability of the 5' nuclease to cleave newly made poly d(A-T) to generate primers that are, in turn, used by the synthetic domain to catalyze the production of even more poly d(A-T). The 5' nuclease is able to cleave poly d(A-T) because poly d(A-T) is self-complementary and easily forms alternate structures at elevated temperatures. These structures are recognized by the 5' nuclease and are then cleaved to generate more primer for the synthesis reaction.

The following is an example of the dual capture assay to detect an antigen(s): A sample to be analyzed for a given antigen(s) is provided. This sample may comprise a mixture of cells; for example, cells infected with viruses display virally-encoded antigens on their surface. If the antigen(s) to be detected are present in solution, they are first attached to a solid support such as the wall of a microtiter dish or to a bead using conventional methodologies. The sample is then mixed with 1) the synthetic domain of a thermostable DNA polymerase conjugated to an antibody which recognizes either a first antigen or a first epitope on an antigen, and 2) the 5' nuclease domain of a thermostable DNA polymerase conjugated to a second antibody which recognizes either a second, distinct antigen or a second epitope on the same antigen as recognized by the antibody conjugated to the synthetic domain. Following an appropriate period to allow the interaction of the antibodies with their cognate antigens (conditions will vary depending upon the antibodies used; appropriate conditions are well known in the art), the sample is then washed to remove unbound antibody-enzyme domain complexes. dATP, dTTP and a small amount of poly d(A-T) is then added to the washed sample and the sample is incubated at elevated temperatures (generally in the range of 60-80°C and more preferably, 70-75°C) to permit the thermostable synthetic and 5' nuclease domains to function. If the sample contains the antigen(s) recognized by both separately conjugated domains of the polymerase, then an exponential increase in poly d(A-T) production occurs. If only the antibody conjugated to the synthetic domain of the polymerase is present in the sample such

that no 5' nuclease domain is present in the washed sample, then only an arithmetic increase in poly d(A-T) is possible. The reaction conditions may be controlled in such a way so that an arithmetic increase in poly d(A-T) is below the threshold of detection. This may be accomplished by controlling the length of time the reaction is allowed to proceed or by adding so little poly d(A-T) to act as template that in the absence of nuclease activity to generate new poly d(A-T) primers very little poly d(A-T) is synthesized.

It is not necessary for both domains of the enzyme to be conjugated to an antibody. One can provide the synthetic domain conjugated to an antibody and provide the 5' nuclease domain in solution or vice versa. In such a case the conjugated antibody-enzyme domain is added to the sample, incubated, then washed. dATP, dTTP, poly d(A-T) and the remaining enzyme domain in solution is then added.

Additionally, the two enzyme domains may be conjugated to oligonucleotides such that target nucleic acid sequences can be detected. The oligonucleotides conjugated to the two different enzyme domains may recognize different regions on the same target nucleic acid strand or may recognize two unrelated target nucleic acids.

The production of poly d(A-T) may be detected in many ways including: 1) use of a radioactive label on either the dATP or dTTP supplied for the synthesis of the poly d(A-T), followed by size separation of the reaction products and autoradiography; 2) use of a fluorescent probe on the dATP and a biotinylated probe on the dTTP supplied for the synthesis of the poly d(A-T), followed by passage of the reaction products over an avidin bead, such as magnetic beads conjugated to avidin; the presence of the florescent probe on the avidin-containing bead indicates that poly d(A-T) has been formed as the fluorescent probe will stick to the avidin bead only if the fluorescenated dATP is incorporated into a covalent linkage with the biotinylated dTTP; and 3) changes fluorescence polarization indicating an increase in size. Other means of detecting the presence of poly d(A-T) include the use of intercalating fluorescence indicators to monitor the increase in duplex DNA formation.

The advantages of the above dual capture assay for detecting antigenic or nucleic acid targets include:

1) No thermocycling of the sample is required. The polymerase domains and the dATP and dTTP are incubated at a fixed temperature (generally about 70°C). After 30 minutes of incubation up to 75% of the added dNTPs are incorporated into poly d(A-T). The lack of thermocycling makes this assay well suited to clinical laboratory settings; there is no need to purchase a thermocycling apparatus and there is no need to maintain very precise temperature control.

2) The reaction conditions are simple. The incubation of the bound enzymatic domains is done in a buffer containing 0.5 mM $MgCl_2$ (higher concentrations may be used), 2-10 mM Tris-Cl, pH 8.5, approximately 50 μM dATP and dTTP. The reaction volume is 10-20 μl and reaction products are detectable within 10-20 minutes.

3) No reaction is detected unless both the synthetic and nuclease activities are present. Thus, a positive result indicates that both probes (antibody or oligonucleotide) have recognized their targets thereby increasing the specificity of recognition by having two different probes bind to the target.

The ability to separate the two enzymatic activities of the DNAP allows for exponential increases in poly d(A-T) production. If a DNAP is used which lacks 5' nuclease activity, such as the Klenow fragment of DNAPEc1, only a linear or arithmetic increase in poly d(A-T) production is possible [Setlow *et al.*, J. Biol. Chem. 247:224 (1972)]. The ability to provide an enzyme having 5' nuclease activity but lacking synthetic activity is made possible by the disclosure of this invention.

## V. Cleavase™ Fragment Length Polymorphism For The Detection Of Secondary Structure

Nucleic acids assume secondary structures which depend on base-pairing for stability. When single strands of nucleic acids (single-stranded DNA, denatured DNA or RNA) with different sequences, even closely related ones, are allowed to fold on themselves, they assume characteristic secondary structures. These differences in

structures account for the ability of single strand conformation polymorphism (SSCP) analysis to distinguish between DNA fragments having closely related sequences.

The 5' nuclease domains of certain DNA polymerases are specific endonucleases that recognize and cleave nucleic acids at specific structures rather than in a sequence-specific manner (as do restriction endonucleases). The isolated nuclease domain of DNAPTaq described herein (termed the enzyme Cleavase™) recognizes the end of a duplex that has non-base paired strands at the ends. The strand with the 5' end is cleaved at the junction between the single strand and the duplex.

Figure 29 depicts a wild-type substrate and a mutant substrate wherein the mutant substrate differs from the wild-type by a single base change (A to G as indicated). According to the method of the present invention, substrate structures form when nucleic acids are denatured and allowed to fold on themselves (See Figure 29, steps 1 and 2). The step of denaturation may be achieved by treating the nucleic acid with heat, low (<3) or high pH (>10), the use of low salt concentrations, the absence of cations, chemicals (*e.g.*, urea, formamide) or proteins (*e.g.*, helicases). Folding or renaturation of the nucleic acid is achieved by lowering of the temperature, addition of salt, neutralization of the pH, withdrawal of the chemicals or proteins.

The manner in which the substrate folds is dependent upon the sequence of the substrate. The 5' nucleases of the invention cleave the structures (See Figure 29, step 3). The end points of the resulting fragments reflect the locations of the cleavage sites. The cleavage itself is dependent upon the formation of a particular structure, not upon a particular sequence at the cleavage site.

When the 5' nucleases of the invention cleave a nucleic acid substrate, a collection of cleavage products or fragments is generated. These fragments constitute a characteristic fingerprint of the nucleic acid which can be detected [*e.g.*, by electrophoresis on a gel (see step 4)]. Changes in the sequence of a nucleic acid (*e.g.*, single point mutation between a wild-type and mutant gene) alter the pattern of cleavage structures formed. When the 5' nucleases of the invention cleave the structures formed by a wild-type and an altered or mutant form of the substrate, the

distribution of the cleavage fragments generated will differ between the two substrates reflecting the difference in the sequence of the two substrates (See Figure 39, step 5).

The Cleavase™ enzyme generates a unique pattern of cleavage products for a substrate nucleic acid. Digestion with the Cleavase™ enzyme can be used to detect single base changes in DNA molecules of great length (*e.g.*, 1.6 kb in length) to produce a characteristic pattern of cleavage products. The method of the invention is termed "Cleavase™ Fragment Length Polymorphism" (CFLP™). However, it is noted that the invention is not limited to the use of the enzyme Cleavase™; suitable enzymatic cleavage activity may be provided from a variety of sources including the Cleavase™ enzyme, *Taq* DNA polymerase, *E. coli* DNA polymerase I and eukaryotic structure-specific endonucleases (*e.g.*, the yeast RAD2 protein and RAD1/RAD10 complex [Harrington, J.J. and Liener (1994) Genes and Develop. 8:1344], murine FEN-1 endonucleases (Harrington and Liener, *supra*) and calf thymus 5' to 3' exonuclease [Murante, R.S., *et al.* (1994) J. Biol. Chem. 269:1191]). Indeed actual experimental data is provided herein which demonstrates that numerous enzymes may be used to generate a unique pattern of cleavage products for a substrate nucleic acid. Enzymes which are shown herein to be suitable for use in the CFLP™ method include the Cleavase™ BN enzyme, *Taq* DNA polymerase, *Tth* DNA polymerase, *Tfl* DNA polymerase, *E. coli* Exo III, and the yeast Rad1/Rad10 complex.

The invention demonstrates that numerous enzymes may be suitable for use in the CFLP™ method including enzymes which have been characterized in the literature a being 3' exonucleases. In order to test whether an enzyme is suitable for use as a cleavage means in the CFLP™ method (*i.e.*, capable of generating a unique pattern of cleavage products for a substrate nucleic acid), the following steps are taken. Careful consideration of the steps described below allows the evaluation of any enzyme ("enzyme X") for use in the CFLP™ method.

An initial CFLP™ reaction is prepared using a previously characterized substrate nucleic acid [for example the 157 nucleotide fragment of exon 4 of the human tyrosinase gene (SEQ ID NO:47)]. The substrate nucleic acid (approximately

100 fmoles; the nucleic acid template may contain a 5' end or other label to permit easy detection of the cleavage products) is placed into a thin wall microcentrifuge tube in a solution which comprises reaction conditions reported to be optimal for the characterized activity of the enzyme (*i.e.*, enzyme X). For example, if the enzyme X is a DNA polymerase, the initial reaction conditions would utilize a buffer which has been reported to be optimal for the polymerization activity of the polymerase. If enzyme X is not a polymerase, or if no specific components are reported to be needed for activity, the initial reaction may be assembled by placing the substrate nucleic acid in a solution comprising 1X CFLP™ buffer (10 mM MOPS, 0.05% Tween-20, 0.05% Nonidet P-40), pH 7.2 to 8.2, 1 mM $MnCl_2$.

The substrate nucleic acid is denatured by heating the sample tube to 95°C for 5 seconds and then the reaction is cooled to a temperature suitable for the enzyme being tested (*e.g.*, if a thermostable polymerase is being tested the cleavage reaction may proceed at elevated temperatures such as 72°C; if a mesophilic enzyme is being tested the tube is cooled to 37°C for the cleavage reaction). Following denaturation and cooling to the target temperature, the cleavage reaction is initiated by the addition of a solution comprising 1 to 200 units of the enzyme to be tested (*i.e.*, enzyme X; the enzyme may be diluted into 1X CFLP™ buffer, pH 8.2 if desired).

Following the addition of the enzyme X solution, the cleavage reaction is allowed to proceed at the target temperature for 2 to 5 minutes. The cleavage reaction is then terminated [this may be accomplished by the addition of a stop solution (95% formamide, 10 mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol)] and the cleavage products are resolved and detected using any suitable method (*e.g.*, electrophoresis on a denaturing polyacrylamide gel followed by transfer to a solid support and nonisotopic detection). The cleavage pattern generated is examined by the criteria described below for the CFLP™ optimization test.

An enzyme is suitable for use in the CFLP™ method if it is capable of generating a unique (*i.e.*, characteristic) pattern of cleavage products from a substrate nucleic acid; this cleavage must be shown to be dependent upon the presence of the

enzyme. Additionally, an enzyme must be able to reproducibly generate the same cleavage pattern when a given substrate is cleaved under the same reaction conditions. To test for reproducibility, the enzyme to be evaluated is used in at least two separate cleavage reactions run on different occasions using the same reaction conditions. If the same cleavage pattern is obtained on both occasions, the enzyme is capable of reproducibly generating a cleavage pattern and is therefore suitable for use in the CFLP™ method.

When enzymes derived from mesophilic organisms are to be tested in the CFLP™ reaction they may be initially tested at 37°C. However it may be desirable to use theses enzymes at higher temperatures in the cleavage reaction. The ability to cleave nucleic acid substrates over a range of temperatures is desirable when the cleavage reaction is being used to detect sequence variation (*i.e.*, mutation) between different substrates. Strong secondary structures that may dominate the cleavage pattern are less likely to be destabilized by single-base changes and may therefore interfere with mutation detection. Elevated temperatures can then be used to bring these persistent structures to the brink of instability, so that the effects of small changes in sequence are maximized and revealed as alterations in the cleavage pattern. Mesophilic enzymes may be used at temperatures greater than 37°C under certain conditions known to the art. These conditions include the use of high (*i.e.*, 10-30%) concentrations of glycerol in the reaction conditions. Furthermore, it is noted that while an enzyme may be isolated from a mesophilic organism this fact alone does not mean that the enzyme may not demonstrate thermostability; therefore when testing the suitability of a mesophilic enzyme in the CFLP™ reaction, the reaction should be run at 37°C and at higher temperatures. Alternatively, mild denaturants can be used to destablize the nucleic acid substrate at a lower temperature (*e.g.*, 1-10% formamide, 1-10% DMSO and 1-10% glycerol have been used in enzymatic reactions to mimic thermal destablization).

Nucleic acid substrates that may be analyzed using a cleavage means, such as a 5' nuclease, include many types of both RNA and DNA. Such nucleic acid substrates may all be obtained using standard molecular biological techniques. For example,

substrates may be isolated from a tissue sample, tissue culture cells, bacteria or viruses, may be transcribed *in vitro* from a DNA template, or may be chemically synthesized. Furthermore, substrates may be isolated from an organism, either as genomic material or as a plasmid or similar extrachromosomal DNA, or it may be a fragment of such material generated by treatment with a restriction endonuclease or other cleavage agents or it may be synthetic.

Substrates may also be produced by amplification using the PCR. When the substrate is to be a single-stranded substrate molecule, the substrate may be produced using the PCR with preferential amplification of one strand (asymmetric PCR). Single-stranded substrates may also be conveniently generated in other ways. For example, a double-stranded molecule containing a biotin label at the end of one of the two strands may be bound to a solid support (*e.g.*, a magnetic bead) linked to a streptavidin moiety. The biotin-labeled strand is selectively captured by binding to the streptavidin-bead complex. It is noted that the subsequent cleavage reaction may be performed using substrate attached to the solid support, as the enzyme Cleavase™ can cleave the substrate while it is bound to the bead. A single-stranded substrate may also be produced from a double-stranded molecule by digestion of one strand with exonuclease.

The nucleic acids of interest may contain a label to aid in their detection following the cleavage reaction. The label may be a radioisotope (*e.g.*, a $^{32}P$ or $^{35}S$-labeled nucleotide) placed at either the 5' or 3' end of the nucleic acid or alternatively the label may be distributed throughout the nucleic acid (*i.e.*, an internally labeled substrate). The label may be a nonisotopic detectable moiety, such as a fluorophore which can be detected directly, or a reactive group which permits specific recognition by a secondary agent. For example, biotinylated nucleic acids may be detected by probing with a streptavidin molecule which is coupled to an indicator (*e.g.*, alkaline phosphatase or a fluorophore), or a hapten such as digoxigenin may be detected using a specific antibody coupled to a similar indicator. Alternatively, unlabeled nucleic acid may be cleaved and visualized by staining (*e.g.*, ethidium bromide staining) or by

hybridization using a labeled probe. In a preferred embodiment, the substrate nucleic acid is labeled at the 5' end with a biotin molecule and is detected using avidin or streptavidin coupled to alkaline phosphatase. In another preferred embodiment the substrate nucleic acid is labeled at the 5' end with a fluorescein molecule and is detected using an anti-fluorescein antibody-alkaline phosphatase conjugate.

The cleavage patterns are essentially partial digests of the substrate in the reaction. When the substrate is labelled at one end (*e.g.*, with biotin), all detectable fragments share a common end. The extension of the time of incubation of the enzyme Cleavase™ reaction does not significantly increase the proportion of short fragments, indicating that each potential cleavage site assumes either an active or inactive conformation and that there is little inter-conversion between the states of any potential site, once they have formed. Nevertheless, many of the structures recognized as active cleavage sites are likely to be only a few base-pairs long and would appear to be unstable at the elevated temperatures used in the Cleavase™ reaction. The formation or disruption of these structures in response to small sequence changes results in changes in the patterns of cleavage.

The products of the cleavage reaction are a collection of fragments generated by structure specific cleavage of the input nucleic acid. Nucleic acids which differ in size may be analyzed and resolved by a number of methods including electrophoresis, chromatography, fluorescence polarization, mass spectrometry and chip hybridization. The invention is illustrated using electrophoretic separation. However, it is noted that the resolution of the cleavage products is not limited to electrophoresis. Electrophoresis is chosen to illustrate the method of the invention because electrophoresis is widely practiced in the art and is easily accessible to the average practitioner.

If abundant quantities of DNA are available for the analysis, it may be advantageous to use direct fluorescence to detect the cleavage fragments, raising the possibility of analyzing several samples in the same tube and on the same gel. This "multiplexing" would permit automated comparisons of closely related substrates such as wild-type and mutant forms of a gene.

The CFLP™ reaction is useful to rapidly screen for differences between similar nucleic acid molecules. To optimize the CFLP™ reaction for any desired nucleic acid system (*e.g.*, a wild-type nucleic acid and one or more mutant forms of the wild-type nucleic acid), it is most convenient to use a single substrate from the test system (for example, the wild-type substrate) to determine the best CFLP™ reaction conditions. A single suitable condition is chosen for doing the comparison CFLP™ reactions on the other molecules of interest. For example, a cleavage reaction may be optimized for a wild-type sequence and mutant sequences may subsequently be cleaved under the same conditions for comparison with the wild-type pattern. The objective of the CFLP™ optimization test is the identification of a set of conditions which allow the test molecule to form an assortment (*i.e.*, a population) of intra-strand structures that are sufficiently stable such that treatment with a structure-specific cleavage agent such as the Cleavase™ enzyme or DNAPTaq will yield a signature array of cleavage products, yet are sufficiently unstable that minor or single-base changes within the test molecule are likely to result in a noticeable change in the array of cleavage products.

The following discussion illustrates the optimization of the CFLP™ method for use with a single-stranded substrate.

A panel of reaction conditions with varying salt concentration and temperature is first performed to identify an optimal set of conditions for the single-stranded CFLP™. "Optimal CFLP™" is defined for this test case as the set of conditions that yields the most widely spaced set of bands after electrophoretic separation, with the most even signal intensity between the bands.

Two elements of the cleavage reaction that significantly affect the stability of the nucleic acid structures are the temperature at which the cleavage reaction is performed and the concentration of salt in the reaction solution. Likewise, other factors affecting nucleic acid structures, such as, formamide, urea or extremes in pH may be used. The initial test typically will comprise reactions performed at four temperatures (60°C, 65°C, 70°C and 75°C) in three different salt concentrations (0 mM, 25 mM and 50 mM) for a total of twelve individual reactions. It is not intended that the present invention be limited by the salt utilized. The salt utilized may be

chosen from potassium chloride, sodium chloride, etc. with potassium chloride being a preferred salt.

For each salt concentration to be tested, 30 µl of a master mix containing a DNA substrate, buffer and salt is prepared. When the substrate is DNA, suitable buffers include 3-[N-Morpholino]propanesulfonic acid (MOPS), pH 6.5 to 9.0, with pH 7.5 to 8.4 being particularly preferred and other "Good" biological buffers such as tris[Hydroxymethyl]aminomethane (Tris) or N,N-bis[2-Hydroxyethyl]glycine (Bicine), pH 6.5 to 9.0, with pH 7.5 to 8.4 being particularly preferred. When the nucleic acid substrate is RNA, the pH of the buffer is reduced to the range of 6.0 to 8.5, with pH 6.0 to 7.0 being particularly preferred. When manganese is to used as the divalent cation in the reaction, the use of Tris buffers is not preferred. Manganese tends to precipitate as manganous oxide in Tris if the divalent cation is exposed to the buffer for prolonged periods (such as in incubations of greater than 5 minutes or in the storage of a stock buffer). When manganese is to be used as the divalent cation, a preferred buffer is the MOPS buffer.

For reactions containing no salt (the "0 mM KCl" mix), the mix includes enough detectable DNA for 5 digests (*e.g.*, approximately 500 fmoles of 5' biotinylated DNA or approximately 100 fmoles of $^{32}$P-5' end labeled DNA) in 30 µl of 1X CFLP™ buffer (10 mM MOPS, pH 8.2) with 1.7 mM $MnCl_2$ or $MgCl_2$ (the final concentration of the divalent cation will be 1 mM). Other concentrations of the divalent cation may be used if appropriate for the cleavage agent chosen (*e.g.*, *E. coli* DNA polymerase I is commonly used in a buffer containing 5 mM $MgCl_2$). The "25 mM KCl" mix includes 41.5 mM KCl in addition to the above components; the "50 mM KCl" mix includes 83.3 mM KCl in addition to the above components.

The mixes are distributed into labeled reaction tubes (0.2 ml, 0.5 ml or 1.5 ml "Eppendorf" style microcentrifuge tubes) in 6 µl aliquots, overlaid with light mineral oil or a similar barrier, and stored on ice until use. Sixty microliters of an enzyme dilution cocktail is assembled, comprising a 5' nuclease at a suitable concentration in 1X CFLP™ buffer without $MnCl_2$. Preferred 5' nucleases and concentrations are 750 ng of the enzyme Cleavase™BN or 15 units of *Taq* DNA polymerase (or another

eubacterial Pol A-type DNA polymerase). Suitable amounts of a similar structure-specific cleavage agent in 1X CFLP™ buffer without $MnCl_2$ may also be utilized.

If a strong (*i.e.*, stable) secondary structure is formed by the substrates, a single nucleotide change is unlikely to significantly alter that structure, or the cleavage pattern it produces. Elevated temperatures can be used to bring structures to the brink of instability, so that the effects of small changes in sequence are maximized, and revealed as alterations in the cleavage pattern within the target substrate, thus allowing the cleavage reaction to occur at that point. Consequently, it is often desirable to run the reaction at an elevated temperature (*i.e.*, above 55°C).

Preferably, reactions are performed at 60°C, 65°C, 70°C and 75°C. For each temperature to be tested, a trio of tubes at each of the three KCl concentrations are brought to 95°C for 5 seconds, then cooled to the selected temperature. The reactions are then started immediately by the addition of 4 µl of the enzyme cocktail. A duplicate trio of tubes may be included (these tubes receiving 4 µl of 1X CFLP™ buffer without enzyme or $MnCl_2$), to assess the nucleic acid stability in these reaction conditions. All reactions proceed for 5 minutes, and are stopped by the addition of 8 µl of 95% formamide with 20 mM EDTA and 0.05% xylene cyanol and 0.05% bromophenol blue. Reactions may be assembled and stored on ice if necessary. Completed reactions are stored on ice until all reactions in the series have been performed.

Samples are heated to 72°C for 2 minutes and 5 µl of each reaction is resolved by electrophoresis through a suitable gel, such as 6 to 10% polyacrylamide (19:1 cross-link), with 7M urea, in a buffer of 45 mM Tris-Borate, pH 8.3, 1.4 mM EDTA for nucleic acids up to approximately 1.5 kb, or native or denaturing agarose gels for larger molecules. The nucleic acids may be visualized as described above, by staining, autoradiography (for radioisotopes) or by transfer to a nylon or other membrane support with subsequent hybridization and/or nonisotopic detection. The patterns generated are examined by the criteria described above and a reaction condition is chosen for the performance of the variant comparison CFLP™s.

A "no enzyme" control allows the assessment of the stability of the nucleic acid substrate under particular reaction conditions. In this instance, the substrate is placed in a tube containing all reaction components except the enzyme and treated the same as the enzyme-containing reactions. Other control reactions may be run. A wild-type substrate may be cleaved each time a new mutant substrate is tested. Alternatively, a previously characterized mutant may be run in parallel with a substrate suspected of containing a different mutation. Previously characterized substrates allow for the comparison of the cleavage pattern produced by the new test substrate with a known cleavage pattern. In this manner, alterations in the new test substrate may be identified.

When the CFLP™ pattern generated by cleavage of a single-stranded substrate contains an overly strong (*i.e.*, intense) band, this indicates the presence of a very stable structure. The preferred method for redistributing the signal is to alter the reaction conditions to increase structure stability (*e.g.*, lower the temperature of the cleavage reaction, raise the monovalent salt concentration); this allows other less stable structures to compete more effectively for cleavage.

When the single-stranded substrate is labelled at one end (*e.g.*, with biotin or 32P) all detectable fragments share a common end. For short DNA substrates (less than 250 nucleotides) the concentration of the enzyme (*e.g.*, Cleavase™ BN) and the length of the incubation have minimal influence on the distribution of signal intensity, indicating that the cleavage patterns are not partial digests of a single structure assumed by the nucleic acid substrate, but rather are relatively complete digests of a collection of stable structures formed by the substrate. With longer DNA substrates (greater than 250 nucleotides) there is a greater chance of having multiple cleavage sites on each structure, giving apparent overdigestion as indicated by the absence of any residual full-length materials. For these DNA substrates, the enzyme concentration may be lowered in the cleavage reaction (for example, if 50 ng of the Cleavase™ BN enzyme were used initially and overdigestion was apparent, the concentration of enzyme may be reduced to 25, 10 or 1 ng per reaction).

When the CFLP™ reaction is to optimized for the cleavage a double-stranded substrate the following steps are taken. The cleavage of double-stranded DNA substrates up to 2,000 base pairs may be optimized in this manner.

The double-stranded substrate is prepared such that it contains a single end-label using any of the methods known to the art. The molar amount of DNA used in the optimization reactions is the same as that use for the optimization of reactions utilizing single-stranded substrates. The most notable differences between the optimization of the CFLP™ reaction for single- versus double-stranded substrates is that the double-stranded substrate is denatured in distilled water without buffer, the concentration of $MnCl_2$ in the reaction is reduced to 0.2 mM, the KCl (or other monovalent salt) is omitted, and the enzyme concentration is reduced to 10 to 25 ng per reaction. In contrast to the optimization of the single-stranded CFLP™ reaction (described above) where the variation of the monovalent salt (*e.g.*, KCl) concentration is a critical controlling factor, in the optimization of the double-stranded CFLP™ reaction.the range of temperature is the more critical controlling factor for optimization of the reaction. When optimizing the double-stranded CFLP™ reaction a reaction tube containing the substrate and other components described below is set up to allow performance of the reaction at each of the following temperatures: 40°C, 45°C, 50°C, 55°C, 60°C, 65°C, 70°C, and 75°C.

For each temperature to be tested, a mixture comprising the single end labelled double-stranded DNA substrate and distilled water in a volume of 15 µl is prepared and placed into a thin walled microcentrifuge tube. This mixture may be overlaid with light mineral oil or liquid wax (this overlay is not generally required but may provide more consistent results with some double-stranded DNA substrates).

A 2 mM solution of $MnCl_2$ is prepared. For each CFLP™ reaction, 5 µl of a diluted enzyme solution is prepared comprising 2 µl of 10X CFLP™ buffer (100 mM MOPS, pH 7.2 to 8.2, 0.5% Tween-20, 0.5% Nonidet P-40), 2 µl of 2 mM $MnCl_2$ and 25 ng of Cleavase™ BN enzyme and distilled water to yield a final volume of 5 µl.

The DNA mixture is heated to 95°C for 10 to 30 seconds and then individual tubes are cooled to the reaction temperatures to be tested (*e.g.*, 40°C, 45°C, 50°C,

55°C, 60°C, 65°C, 70°C, and 75°C). The cleavage reaction is started by adding 5 µl of the dilute enzyme solution to each tube at the target reaction temperature. The reaction is incubated at the target temperature for 5 minutes and the reaction is terminated (*e.g.*, by the addition of 16 µl of stop solution comprising 95% formamide with 10 mM EDTA and 0.05% xylene cyanol and 0.05% bromophenol blue).

Samples are heated to 72°C for 1 to 2 minutes and 3 to 7 µl of each reaction is resolved by electrophoresis through a suitable gel, such as 6 to 10% polyacrylamide (19:1 cross-link), with 7M urea, in a buffer of 45 mM Tris-Borate, pH 8.3, 1.4 mM EDTA for nucleic acids up to approximately 1.5 kb, or native or denaturing agarose gels for larger molecules. The nucleic acids may be visualized as described above, by staining, autoradiography (for radioisotopes) or by transfer to a nylon or other membrane support with subsequent hybridization and/or nonisotopic detection. The patterns generated are examined by the criteria described above and a reaction condition is chosen for the performance of the double-stranded CFLP™.

A "no enzyme" control allows the assessment of the stability of the nucleic acid substrate under particular reaction conditions. In this instance, the substrate is placed in a tube containing all reaction components except the enzyme and treated the same as the enzyme-containing reactions. Other control reactions may be run. A wild-type substrate may be cleaved each time a new mutant substrate is tested. Alternatively, a previously characterized mutant may be run in parallel with a substrate suspected of containing a different mutation. Previously characterized substrates allow for the comparison of the cleavage pattern produced by the new test substrate with a known cleavage pattern. In this manner, alterations in the new test substrate may be identified.

When performing double-stranded CFLP™ reactions the $MnCl_2$ concentration preferably will not exceed 0.25 mM. If the end label on the double-stranded DNA substrate disappears (*i.e.*, loses its 5' end label as judged by a loss of signal upon detection of the cleavage products), the concentration of $MnCl_2$ may be reduced to 0.1 mM. Any EDTA present in the DNA storage buffer will reduce the amount of free

$Mn^{2+}$ in the reaction, so double-stranded DNA should be dissolved in water or Tris-HCl with a EDTA concentration of 0.1 mM or less.

Cleavage products produced by cleavage of either single-or double-stranded substrates which contain a biotin label may be detected using the following nonisotopic detection method. After electrophoresis of the reaction products, the gel plates are separated allowing the gel to remain flat on one plate. A positively charged nylon membrane (preferred membranes include Nytran®Plus, 0.2 or 0.45 mm-pore size, Schleicher and Schuell, Keene, NH), cut to size and pre-wetted in 0.5X TBE (45 mM tris-Borate, pH 8.3, 1.4 mM EDTA), is laid on top of the exposed gel. All air bubbles trapped between the gel and the membrane are removed (*e.g.*, by rolling a 10 ml pipet firmly across the membrane). Two pieces of 3MM filter paper (Whatman) are then placed on top of the membrane, the other glass plate is replaced, and the sandwich is clamped with binder clips or pressed with books or weights. The transfer is allowed to proceed 2 hours to overnight (the signal increases with longer transfer).

After transfer, the membrane is carefully peeled from the gel and allowed to air dry. Distilled water from a squeeze bottle can be used to loosen any gel that sticks to the membrane. After complete drying, the membrane is agitated for 30 minutes in 1.2X Sequenase Images Blocking Buffer (United States Biochemical, Cleveland, OH; avoid any precipitates in the blocking buffer by decanting or filtering); 0.3 ml of the buffer is used per $cm^2$ of membrane (*e.g.*, 30 mls for a 10cm x 10cm blot). A streptavidin-alkaline phosphatase conjugate (SAAP, United Stated Biochemical) is added at a 1:4000 dilution directly to the blocking solution (avoid spotting directly on membrane), and agitated for 15 minutes. The membrane is rinsed briefly with $dH_2O$ and then washed 3 times (5 minutes of shaking per/wash) in 1X SAAP buffer (100 mM Tris-HCl, pH 10; 50 mM NaCl) with 0.1% sodium dodecyl sulfate (SDS), using 0.5 ml buffer/$cm^2$ of membrane, with brief water rinses between each wash. The membrane is then washed twice in 1X SAAP buffer (no SDS) with 1 mM $MgCl_2$, drained thoroughly, and placed in a plastic heat-sealable bag. Using a sterile pipet tip, 0.05 ml/$cm^2$ of CDP-Star™ (Tropix, Bedford, MA) is added to the bag and distributed over the entire membrane for 5 minutes. The bag is drained of all excess liquid and

air bubbles, sealed, and the membrane is exposed to X-ray film (*e.g.*, Kodak XRP) for 30 minutes. Exposure times are adjusted as necessary for resolution and clarity.

To date, every nucleic acid substrate tested in the CFLP™ system has produced a reproducible pattern of fragments. The sensitivity and specificity of the cleavage reaction make this method of analysis very suitable for the rapid screening of mutations in cancer diagnostics, tissue typing, genetic identity, bacterial and viral typing, polymorphism analysis, structure analysis, mutant screening in genetic crosses, etc. It could also be applied to enhanced RNA analysis, high level multiplexing and extension to longer fragments. One distinct benefit of using the Cleavase™ reaction to characterize nucleic acids is that the pattern of cleavage products constitutes a characteristic fingerprint, so a potential mutant can be compared to previously characterized mutants without sequencing. Also, the place in the fragment pattern where a change is observed gives a good indication of the position of the mutation. But it is noted that the mutation need not be at the precise site of cleavage, but only in an area that affects the stability of the structure.

## VI. Detection of Mutations in the p53 Tumor Suppressor Gene Using the CFLP™ Method

Tumor supressor genes control cellular proliferation and a variety of other processes important for tissue homeostasis. One of the most extensively studied of these, the p53 gene, encodes a regulator of the cell cycle machinery that can suppress the growth of cancer cells as well as inhibit cell transformation (Levine, Annu. Rev. Biochem. 62:623 [1993]). Tumor supressor mutations that alter or obliterate normal p53 function are common.

Mutations in the p53 tumor supressor gene are found in about half of all cases of human cancer making alterations in the p53 gene the most common cancer-related genetic change known at the gene level. In the wild-type or non-mutated form, the p53 gene encodes a 53-kD nuclear phosphoprotein, comprising 393 amino acids, which is involved in the control of cellular proliferation. Mutations in the p53 gene are generally (greater than 90%) missense mutations which cause a change in the identity

of an amino acid rather than nonsense mutations which cause inactivation of the protein. It has been postulated that the high frequency of p53 mutation seen in human tumors is due to the fact that the missense mutations cause both a loss of tumor supressor function and a gain of oncogenic function [Lane, D.P. and Benchimol, S., Genes Dev. 4:1 (1990)].

The gene encoding the p53 protein is large, spanning 20,000 base pairs, and is divided into 11 exons (see Figure 76). The ability to scan the large p53 gene for the presence of mutations has important clinical applications. In several major human cancers the presence of a tumor p53 mutation is associated with a poor prognosis. p53 mutation has been shown to be an independent marker of reduced survival in lymph node-negative breast cancers, a finding that may assist clinicians in reaching decisions regarding more aggressive therapeutic treatment. Also, Lowe and co-workers have demonstrated that the vulnerability of tumor cells to radiation or chemotherapy is greatly reduced by mutations which abolish p53-dependent apoptosis [Lowe *et al.*, Cell 74:957 (1995)].

Regions of the p53 gene from approximately 10,000 tumors have been sequenced in the last 4 to 5 years, resulting in characterization of over 3,700 mutations of which approximately 1,200 represent independent p53 mutations (*i.e.*, point mutations, insertion or deletions). A database has been compiled and deposited with the European Molecular Biology Laboratory (EMBL) Data Library and is available in electronic form [Hollstein, M. *et al.* (1994) Nucleic Acids Res. 22:3551 and Cariello, N.F. *et al.* (1994) Nucleic Acids Res. 22:3549]. In addition, an IBM PC compatible software package to analyze the information in the database has been developed. [Cariello *et al.*, Nucel. Acids Res. 22:3551 (1994)]. The point mutations in the database were identified by DNA sequencing of PCR-amplified products. In most cases, preliminary screening for mutations by SSCP or DGGE was performed.

Analysis of the p53 mutations shows that the p53 gene contains 5 hot spot regions (HSR) most frequently mutated in human tumors that show a tight correlation between domains of the protein that are evolutionary highly conserved (ECDs) and

seem to be specifically involved in the transformation process (see Figure 76; the hieght of the bar represent the relative percentage of total mutations associated with the five HSRs). The five HSRs are confined to exons 5 to 8 and account for over 85% of the mutations detected. However, becuase these studies generally confined their analysis to PCR amplifications and sequencing of regions located between exons 5 to 8, it should be kept in mind that mutations outside this region are underrepresented. As 10% to 15% of the mutations lie outside this region, a clinically effective p53 gene DNA diagnostic should be able to cost-effectively scan for life-threatening mutations scattered across the entire gene (33).

The following table lists a number of the known p53 mutations.



**HUMAN p53 GENE MUTATIONS**

TABLE 2

| CODON NO. | WILD-TYPE | MUTANT | EVENT | TUMOR TYPE |
|---|---|---|---|---|
| 36 | CCG | CCA | GC→AT | Lung |
| 49 | GAT | CAT | GC→CG | CML |
| 53 | TGG | TGT | GC→TA | CML |
| 60 | CCA | TCA | GC→AT | CML |
| 68 | GAG | TAG | GC→TA | SCLC |
| 110 | CGT | TGT | GC→AT | Hepatoca |
| 113 | TTC | TGT | Double M | NSCLC |
| 128 | CCT | CCG | T→G | Breast |
| 128 | | TCT | C→T | Breast |
| 129 | GCC | GAC | GC→TA | Neurofibrosa |
| 130 | CTC | CTG | GC→CG | MDS |
| 132 | AAG | AAC | GC→CG | Colorectal ca |
| 132 | | CAG | AT→CG | Breast ca |
| 132 | | AAT | GC→TA | Lung (NSCLC) ca |
| 132 | | CAG | AT→CG | Pancreatic ca |
| 132 | | AGG | AT→GC | CML |

| | | | | |
|---|---|---|---|---|
| 133 | ATG | TTG | AT→TA | Colorectal ca |
| 133 | | AAG | AT→TA | Burkitt lymphoma |
| 134 | TTT | TTA | AT→TA | Lung (SCLC) ca |
| 135 | TGC | TAC | GC→AT | Colorectal ca |
| 135 | | TCC | GC→CG | AML |
| 135 | | TAC | GC→AT | Lung (NSCLC) ca |
| 135 | | TGG | GC→CG | MDS |
| 136 | CAA | GAG | Double M | Breast ca |
| 138 | GCC | GTC | GC→AT | Rhabdomyosa |
| 138 | | GGC | GC→CG | Lung (SCLC) ca |
| 140 | ACC | TAC | AT→TA | CML |
| 141 | TGC | TAC | GC→AT | Colorectal ca |
| 141 | | TAC | GC→AT | Bladder ca |
| 143 | GTG | GCG | AT→GC | Colorectal ca |
| 143 | | TTG | GC→TA | Lung (NSCLC) ca |
| 144 | CAG | TAG | GC→AT | Esophageal ca |
| 144 | | CCG | AT→CG | Burkitt lymphoma |
| 151 | CCC | CAT | Double M | Leiomyosa |
| 151 | | CAC | GC→TA | Lung (SCLC) ca |
| 151 | | TCC | GC→AT | Glioblastoma |
| 151 | | TCC | GC→AT | Lung (NSCLC) ca |
| 152 | CCG | CTG | GC→AT | Leiomyosa |
| 152 | | TCG | GC→AT | Breast ca |
| 154 | GGC | GTC | GC→TA | Esophageal ca |
| 154 | | GTC | GC→TA | Lung (NSCLC) ca |
| 154 | | GTC | GC→TA | Lung (NSCLC) ca |
| 154 | | GTC | GC→TA | Lung (NSCLC) ca |
| 156 | CGC | CCC | GC→CG | Rhabdomyosa |
| 156 | | CCC | GC→CG | Osteosa |
| 156 | | CGT | GC→AT | Lung (NSCLC) ca |
| 156 | | CCC | GC→CG | Lung (NSCLC) ca |
| 157 | GTC | TTC | GC→TA | Hepatoca |

| | | | | |
|---|---|---|---|---|
| 157 | | TTC | GC→TA | Lung (SCLC) ca |
| 157 | | TTC | GC→TA | Lung (NSCLC) ca |
| 157 | | TTC | GC→TA | Breast ca |
| 157 | | TTC | GC→TA | Lung (SCLC) ca |
| 157 | | TTC | GC→TA | Bladder ca |
| 158 | CGC | CGT | GC→AT | Neurofibrosa |
| 158 | | CAC | GC→AT | Burkitt lymphoma |
| 159 | GCC | GTC | GC→AT | Lung (NSCLC) ca |
| 159 | | CCC | GC→CG | Lung (NSCLC) ca |
| 163 | TAC | TGC | AT→GC | Breast ca |
| 163 | | CAC | AT→GC | Burkitt lymphoma |
| 164 | AAG | CAG | AT→CG | Breast ca |
| 171 | GAG | TAG | GC→TA | Lung (SCLC) ca |
| 172 | GTT | TTT | GC→TA | Burkitt lymphoma |
| 173 | GTG | TTG | GC→TA | Lung (NSCLC) ca' |
| 173 | | TTG | GC→TA | Lung (NSCLC) ca |
| 173 | | GGG | AT→CG | Burkitt lymphoma |
| 173 | | GTA | GC→AT | Gastric ca |
| 175 | CGC | CAC | GC→AT | Colorectal ad |
| 175 | | CAC | GC→AT | Colorectal ad |
| 175 | | CAC | GC→AT | Colorectal ad |
| 175 | | CAC | GC→AT | Colorectal ca |
| 175 | | CAC | GC→AT | Colorectal ca |
| 175 | | CAC | GC→AT | T-ALL |
| 175 | | CAC | GC→AT | Brain tumor |
| 175 | | CAC | GC→AT | Colorectal ca |
| 175 | | CAC | GC→AT | Colorectal ca |
| 175 | | CAC | GC→AT | Leiomyosa |
| 175 | | CAC | GC→AT | Esophageal ca |
| 175 | | CAC | GC→AT | Glioblastoma |
| 175 | | CAC | GC→AT | Colorectal ca |
| 175 | | CAC | GC→AT | T-ALL |

| | | | | |
|---|---|---|---|---|
| 175 | | CAC | GC→AT | Breast ca |
| 175 | | CTC | GC→TA | Breast ca |
| 175 | | AGC | GC→TA | Hepatoca |
| 175 | | CAC | GC→AT | B-ALL |
| 175 | | CAC | GC→AT | B-ALL |
| 175 | | CAC | GC→AT | Burkitt lymphoma |
| 175 | | CAC | GC→AT | Burkitt lymphoma |
| 175 | | CAC | GC→AT | Burkitt lymphoma |
| 175 | | CAC | GC→AT | Burkitt lymphoma |
| 175 | | CAC | GC→AT | Gastric ca |
| 176 | TGC | TTC | GC→TA | Lung (NSCLC) ca |
| 176 | | TTC | GC→TA | Esophageal ca |
| 176 | | TTC | GC→TA | Lung (NSCLC) ca |
| 176 | | TAC | GC→AT | Burkitt lymphoma |
| 177 | CCC | CGC | GC→CG | PTLC |
| 179 | CAT | TAT | GC→AT | Neurofibrosa |
| 179 | | CAG | AT→CG | Lung (SCLC) ca |
| 179 | | CTT | AT→TA | Esophageal ca |
| 179 | | GAT | GC→CG | Breast ca |
| 179 | | CTT | AT→TA | Cholangiosa |
| 179 | | CTT | AT→TA | Cholangiosa |
| 181 | CGC | CAC | GC→AT | Li-Fraumeni sdm |
| 187 | GGT | TGT | GC→TA | Breast ca |
| 192 | CAG | TAG | GC→AT | Esophageal ca |
| 193 | CAT | CGT | AT→GC | Lung (SCLC) ca |
| 193 | | TAT | GC→AT | Esophageal ca |
| 193 | | CGT | AT→GC | AML |
| 194 | CTT | TTT | GC→AT | Breast ca |
| 194 | | CGT | AT→CG | Lung (SCLC) ca |
| 194 | | CGT | AT→CG | Esophageal ca |
| 194 | | CGT | AT→CG | Esophageal ca |
| 194 | | CGT | AT→CG | B-CLL |

| | | | | |
|---|---|---|---|---|
| 196 | CGA | TGA | GC→AT | Colorectal ca |
| 196 | | TGA | GC→AT | T-ALL |
| 196 | | TGA | GC→AT | T-cell lymphoma |
| 196 | | TGA | GC→AT | Lung (SCLC) ca |
| 196 | | TGA | GC→AT | Bladder ca |
| 198 | GAA | TAA | GC→TA | Lung (SCLC) ca |
| 198 | | TAA | GC→TA | Lung (SCLC) ca |
| 202 | CGT | CTT | GC→TA | CML |
| 204 | GAG | GGG | AT→GC | CML |
| 205 | TAT | TGT | AT→GC | B-ALL |
| 205 | | TGT | AT→GC | B-CLL |
| 205 | | TTT | AT→TA | Gastric ca |
| 211 | ACT | GCT | AT→GC | Colorectal ca |
| 213 | CGA | TGA | GC→AT | Colorectal ca |
| 213 | | CAA | GC→AT | B-cell lymphoma |
| 213 | | CAA | GC→AT | Burkitt lymphoma |
| 213 | | CGG | AT→GC | Lung (SCLC) ca |
| 213 | | CGG | AT→GC | Esophageal ca |
| 213 | | TGA | GC→AT | Lung (NSCLC) ca |
| 213 | | CGG | AT→GC | Lung (NSCLC) ca |
| 213 | | TGA | GC→AT | Burkitt lymphoma |
| 213 | | TGA | GC→AT | Burkitt lymphoma |
| 215 | AGT | GGT | AT→GC | Colorectal ca |
| 216 | GTG | ATG | GC→AT | Brain tumor |
| 216 | | GAG | AT→TA | Burkitt lymphoma |
| 216 | | TTG | GC→TA | Gastric ca |
| 216 | | ATG | GC→AT | Ovarian ca |
| 220 | TAT | TGT | AT→GC | Colorectal ca |
| 229 | TGT | TGA | AT→TA | Lung (SCLC) ca |
| 232 | ATC | AGC | AT→CG | B-CLL |
| 234 | TAC | CAC | AT→GC | B-cell lymphoma |
| 234 | | CAC | AT→GC | Burkitt lymphoma |

| | | | | |
|---|---|---|---|---|
| 234 | | TGC | AT→GC | Burkitt lymphoma |
| 236 | TAC | TGC | AT→GC | Burkitt lymphoma |
| 237 | ATG | AGG | AT→CG | T-ALL |
| 237 | | ATA | GC→AT | Lung (SCLC) ca |
| 237 | | ATA | GC→AT | AML |
| 237 | | ATA | GC→AT | Breast ca |
| 237 | | ATA | GC→AT | Burkitt lymphoma |
| 237 | | ATA | GC→AT | Richter's sdm |
| 238 | TGT | TTT | GC→TA | Larynx ca |
| 238 | | TAT | GC→AT | Burkitt lymphoma |
| 238 | | TAT | GC→AT | CML |
| 239 | AAC | AGC | AT→GC | Colorectal ca |
| 239 | | AGC | AT→GC | Colorectal ca |
| 239 | | AGC | AT→GC | Burkitt lymphoma |
| 239 | | AGC | AT→GC | CML |
| 239 | | AGC | AT→GC | CML |
| 239 | | AGC | AT→GC | B-CLL |
| 241 | TCC | TTC | GC→AT | Colorectal ca |
| 241 | | TGC | GC→CG | Colorectal ca |
| 241 | | TGC | GC→CG | Bladder ca |
| 242 | TGC | TCC | GC→CG | Lung (SCLC) ca |
| 242 | | TTC | GC→TA | Breast ca |
| 242 | | TCC | GC→CG | MDS |
| 242 | | TAC | GC→AT | Ependymoma |
| 244 | GGC | TGC | GC→TA | T-ALL |
| 244 | | TGC | GC→TA | Esphageal ca |
| 244 | | TGC | GC→TA | Lung (SCLC) ca |
| 244 | | AGC | GC→AT | Hepatoca |
| 245 | GGC | GTC | GC→TA | Esophageal ca |
| 245 | | TGC | GC→TA | Li-Fraumeni sdm |
| 245 | | AGC | GC→AT | Leyomyosa |
| 245 | | GAC | GC→AT | Li-Fraumeni sdm |

| | | | | |
|---|---|---|---|---|
| 245 | | AGC | GC→AT | Esophageal ca |
| 245 | | GCC | GC→CG | Bladder ca |
| 245 | | GAC | GC→AT | Breast ca |
| 245 | | GAC | GC→AT | Li-Fraumeni sdm |
| 245 | GGC | TGC | GC→TA | Li-Fraumeni sdm |
| 245 | | GTC | GC→TA | Cervical ca |
| 246 | ATG | GTG | AT→GC | AML |
| 246 | | ATC | GC→CG | Lung (NSCLC) ca |
| 246 | | GTG | AT→GC | Hepatoca |
| 246 | | GTG | AT→GC | Bladder ca |
| 247 | AAC | ATC | AT→TA | Lung (NSCLC) ca |
| 248 | CGG | TGG | GC→AT | Colorectal ad |
| 248 | | TGG | GC→AT | Colorectal ca |
| 248 | | CAG | GC→AT | Colorectal ca |
| 248 | | CAG | GC→AT | Colorectal ca |
| 248 | | CAG | GC→AT | T-ALL |
| 248 | | CAG | GC→AT | Esophageal ca |
| 248 | | TGG | GC→AT | Li-Fraumeni sdm |
| 248 | | TGG | GC→AT | Li-Fraumeni sdm |
| 248 | | TGG | GC→AT | Colorectal ca |
| 248 | | TGG | GC→AT | Colorectal ca |
| 248 | | TGG | GC→AT | Rhabdomyosa |
| 248 | | CTG | GC→TA | Esophageal ca |
| 248 | | TGG | GC→AT | Lung (NSCLC) ca |
| 248 | | CAG | GC→AT | Lung (SCLC) ca |
| 248 | | CTG | GC→TA | Lung (SCLC) ca |
| 248 | | CAG | GC→AT | T-ALL |
| 248 | | TGG | GC→AT | Lung (NSCLC) ca |
| 248 | | CTG | GC→TA | Lung (SCLC) ca |
| 248 | | TGG | GC→AT | Colorectal ca |
| 248 | | CAG | GC→AT | Bladder ca |
| 248 | | CAG | GC→AT | MDS |

| | | | | |
|---|---|---|---|---|
| 248 | | TGG | GC→AT | Burkitt lymphoma |
| 248 | | CAG | GC→AT | Breast ca |
| 248 | | CAG | GC→AT | B-CLL |
| 248 | | CAG | GC→AT | Burkitt lymphoma |
| 248 | | TGG | GC→AT | Burkitt lymphoma |
| 248 | | CAG | GC→AT | Burkitt lymphoma |
| 248 | | TGG | GC→AT | Burkitt lymphoma |
| 248 | | CAG | GC→AT | Gastric ca |
| 248 | | TGG | GC→AT | Lung (SCLC) ca |
| 248 | | CAG | GC→AT | Breast ca |
| 248 | | CAG | GC→AT | CML |
| 248 | | TGG | GC→AT | Li-Fraumeni sdm |
| 248 | | CAG | GC→AT | Li-Fraumeni sdm |
| 248 | | TGG | GC→AT | Colorectal ca |
| 249 | AGG | AGT | GC→TA | Hepatoca |
| 249 | | AGT | GC→TA | Hepatoca |
| 249 | | AGT | GC→TA | Hepatoca |
| 249 | | AGC | GC→CG | Hepatoca |
| 249 | | AGT | GC→TA | Hepatoca |
| 249 | | AGT | GC→TA | Hepatoca |
| 249 | | AGT | GC→TA | Hepatoca |
| 249 | | AGT | GC→TA | Hepatoca |
| 249 | | AGT | GC→TA | Hepatoca |
| 249 | | AGT | GC→TA | Hepatoca |
| 249 | | AGT | GC→TA | Esophageal ca |
| 249 | | AGC | GC→CG | Breast ca |
| 249 | | AGT | GC→TA | Lung (NSCLC) ca |
| 249 | | AGT | GC→TA | Hepatoca |
| 250 | CCC | CTC | GC→AT | Burkitt lymphoma |
| 251 | ATC | AGC | AT→CG | Gastric ca |
| 252 | CTC | CCC | AT→GC | Li-Fraumeni sdm |

| | | | | |
|---|---|---|---|---|
| 252 | CTC | CCC | AT→GC | Li-Fraumeni sdm |
| 254 | ATC | GAC | Double M | Burkitt lymphoma |
| 254 | | AAC | AT→TA | Breast ca |
| 256 | ACA | GCA | AT→GC | T-ALL |
| 258 | GAA | AAA | GC→AT | Li-Fraumeni sdm |
| 258 | | AAA | GC→AT | Burkitt lymphoma |
| 258 | | AAA | GC→AT | Li-Fraumeni sdm |
| 259 | GAC | GGC | AT→GC | T-ALL |
| 260 | TCC | GCC | AT→CG | T-ALL |
| 266 | GGA | GTA | GC→TA | Lung (NSCLC) ca |
| 266 | | GTA | GC→TA | Lung (NSCLC) ca |
| 266 | | GTA | GC→TA | Breast ca |
| 267 | CGG | CCG | GC→CG | Lung (SCLC) ca |
| 270 | TTT | TGT | AT→CG | Esophageal ca |
| 270 | | TGT | AT→CG | T-ALL |
| 272 | GTG | ATG | GC→AT | Brain tumor |
| 272 | | CTG | GC→CG | Lung (SCLC) ca |
| 272 | | ATG | GC→AT | Hepatoca |
| 272 | | ATG | GC→AT | AML |
| 273 | CGT | TGT | GC→AT | Colorectal ad |
| 273 | | TGT | GC→AT | Brain tumor |
| 273 | | CAT | GC→AT | Breast ca |
| 273 | | CAT | GC→AT | Colorectal ca |
| 273 | | TGT | GC→AT | Lung (NSCLC) ca |
| 273 | | CTT | GC→TA | Lung (SCLC) ca |
| 273 | | CAT | GC→AT | Colorectal ca |
| 273 | | CAT | GC→AT | Colorectal ca |
| 273 | | CAT | GC→AT | Colorectal ca |
| 273 | | CAT | GC→AT | Lung (NSCLC) ca |
| 273 | | CCT | GC→CG | Lung (NSCLC) ca |
| 273 | | CTT | GC→TA | Lung (NSCLC) ca |
| 273 | | CTT | GC→TA | Lung (NSCLC) ca |

| | | | | |
|---|---|---|---|---|
| 273 | | CAT | GC→AT | Thyroid ca |
| 273 | | CAT | GC→AT | Lung (SCLC) ca |
| 273 | | TGT | GC→AT | B-cell lymphoma |
| 273 | | TGT | GC→AT | B-ALL |
| 273 | | TGT | GC→AT | Burkitt lymphoma |
| 273 | | TGT | GC→AT | Burkitt lymphoma |
| 273 | | CAT | GC→AT | Li-Fraumeni sdm |
| 273 | | TGT | GC→AT | Cervical ca |
| 273 | | TGT | GC→AT | AML |
| 273 | | CAT | GC→AT | B→CLL |
| 273 | | CTT | GC→TA | B-CLL |
| 274 | GTT | GAT | AT→TA | Erythroleukemia |
| 276 | GCC | CCC | GC→CG | B-ALL |
| 276 | | GAC | GC→TA | Hepatoca |
| 277 | TGT | TTT | GC→TA | Lung (SCLC) ca |
| 278 | CCT | TCT | GC→AT | Esophageal ca |
| 278 | | CTT | GC→AT | Esophageal ca |
| 278 | | GCT | GC→CG | Breast ca |
| 278 | | TCT | GC→AT | Lung (SCLC) ca |
| 278 | | CGT | GC→CG | Ovarian ca |
| 280 | AGA | AAA | GC→AT | Esophageal ca |
| 280 | | AAA | GC→AT | Breast ca |
| 281 | GAC | GGC | AT→GC | Colorectal ca |
| 281 | | GGC | AT→GC | Breast ca |
| 281 | GAC | GAG | GC→CG | Richter's sdm |
| 281 | | TAC | GC→TA | B-CLL |
| 282 | CGG | TGG | GC→AT | Colorectal ad |
| 282 | | TGG | GC→AT | Colorectal ca |
| 282 | CGG | TGG | GC→AT | Rhabdomyosa |
| 282 | | GGG | GC→CG | Lung (NSCLC) ca |
| 282 | | CCG | GC→CG | Breast ca |
| 282 | | TGG | GC→AT | Bladder ca |

| | | | | |
|---|---|---|---|---|
| 282 | | TGG | GC→AT | AML |
| 282 | | CTG | GC→TA | Breast ca |
| 282 | | TGG | GC→AT | B-ALL |
| 282 | | TGG | GC→AT | Burkitt lymphoma |
| 282 | | TGG | GC→AT | Richter's sdm |
| 282 | | TGG | GC→AT | Ovarian ca |
| 282 | | TGG | GC→AT | Li-Fraumeni sdm |
| 283 | CGC | TGC | GC→AT | Colorectal ca |
| 283 | | CCC | GC→CG | Lung (NSCLC) ca |
| 285 | GAG | AAG | GC→AT | Breast ca |
| 286 | GAA | AAA | GC→AT | Colorectal ca |
| 286 | | GGA | AT→GC | Lung (SCLC) ca |
| 286 | | GCA | AT→CG | Li-Fraumeni sdm |
| 287 | GAG | TAG | GC→TA | Burkitt lymphoma |
| 293 | GGG | TGG | GC→TA | Glioblastoma |
| 298 | GAG | TAG | GC→TA | Bladder ca |
| 302 | GGG | GGT | GC→TA | Lung (SCLC) ca |
| 305 | AAG | TAG | AT→TA | Esophageal ca |
| 305 | | TAG | AT→TA | Esophageal ca |
| 307 | GCA | ACA | GC→AT | Breast ca |
| 309 | CCC | TCC | GC→AT | Colorectal ca |
| 334 | GGG | GTG | GC→TA | Lung (SCLC) ca |
| 342 | CGA | TGA | GC→AT | Lung (SCLC) ca |

DELETIONS/INSERTIONS

| CODON | EVENT | TUMOR TYPE |
|---|---|---|
| 137 | del 7 | Gastric ca |
| 143 | del 1 | Gastric ca |
| 152 | del 13 | Colorectal ad |
| 167 | del 1 | Breast ca |
| 168 | del 31 | Hepatoca |
| 175 | del 18 | Breast ca |
| 190 | del 3 | nul ALL |
| 201 | del 1 | Breast ca |
| 206 | del 1 | Burkitt lymphoma |
| 206 | del 1 | Burkitt lymphoma |
| 214 | del 1 | B-ALL |
| 236 | del 27 | Bladder ca |
| 239 | del 1 | Lung (NSCLC) ca |
| 262 | del 1 | Astrocytoma |
| 262 | del 24 | Gastric ca |
| 262 | del 24 | Lung (NSCLC) ca |
| 263 | del 1 | Esophageal ca |
| 264 | del 1 | AML |
| 286 | del 8 | Hepatoca |
| 293 | del 1 | Lung (NSCLC) ca |
| 307 | del 1 | Li-Fraumeni sdm |
| 381 | del 1 | Hepatoca |
| Exon 5 | del 15 | B-ALL |
| 152 | ins 1 | B-CLL |
| 239 | ins 1 | Waldenstrom sdm |
| 252 | ins 4 | Gastric ca |
| 256 | ins 1 | AML |
| 275 | ins 1 | B-CLL |
| 301 | ins 1 | MDS |
| 307 | ins 1 | Glioblastoma |
| Exon 8 | ins 25 | HCL |

SPLICE MUTATIONS

| INTRON | SITE | EVENT | TUMOR TYPE |
|---|---|---|---|
| Intron 3 | Accept | GC→CG | Lung (SCLC) ca |
| Intron 4 | Donor | GC→TA | Lung (SCLC) ca |
| Intron 4 | Donor | GC→AT | T→ALL |
| Intron 5 | Donor | GC→AT | CML |
| Intron 6 | Donor | AT→CG | Lung (SCLC) ca |
| Intron 6 | Accept | AT→TA | Lung (SCLC) ca |
| Intron 6 | Accept | AT→TA | Lung (NSCLC) ca |
| Intron 7 | Donor | GC→TA | Lung (NSCLC) ca |
| Intron 7 | Accept | GC→CG | Lung (SCLC) ca |
| Intron 7 | Accept | CG→AT | AML |
| Intron 7 | Donor | GC→TA | Lung (SCLC) ca |
| Intron 9 | Donor | GC→TA | Lung (SCLC) ca |

A. **CFLP™ Analysis of p53 Mutations in Clinical Samples**

To permit the identiifcation of mutations in the p53 gene from clinical samples, nucleic acid comprising p53 gene sequences are prepared. The nucleic acid may comprise genomic DNA, RNA or cDNA forms of the p53 gene. Nucleic acid may be extracted from a variety of clinical samples [fresh or frozen tissue, suspensions of cells (*e.g.*, blood), cerebral spinal fluid, sputum, etc.] using a variety of standard techniques or commerically available kits. For example, kits which allow the isolation of RNA or DNA from tissue samples are available from Qiagen, Inc. (Chatsworth, CA) and Stratagene (LaJolla, CA), respectively. Total RNA may be isolated from tissues and tumors by a number of methods known to those skilled in the art and commercial kits are available to facilitate the isolation. For example, the RNeasy® kit (Qiagen Inc., Chatsworth, CA) provides protocol, reagents and plasticware to permit the isolation of total RNA from tissues, cultured cells or bacteria, with no modification to the manufacturer's instructions, in approximately 20 minutes. Should it be desirable, in the case of eukaryotic RNA isolates, to further enrich for messenger RNAs, the

polyadenylated RNAs in the mixture may be specifically isolated by binding to an oligo-deoxythymidine matrix, through the use of a kit such as the Oligotex® kit (Qiagen). Comparable isolation kits for both of these steps are available through a number of commercial suppliers.

In addition, RNA may be extracted from samples, including biopsy specimens, convienently by lysing the homogenized tissue in a buffer containing 0.22 M NaCl, 0.75 mM $MgCl_2$, 0.1 M Tris-HCl, pH 8.0, 12.5 mM EDTA, 0.25% NP40, 1% SDS, 0.5 mM DTT, 500 u/ml placental RNAse inhibitor and 200 µg/ml Proteinase K. Following incubation at 37°C for 30 min, the RNA is extracted with phenol:chloroform (1:1) and the RNA is recovered by ethanol precipitation.

Since the majority of p53 mutations are found within exons 5-8, it is convenient as a first analysis to examine a PCR fragment spanning this region. PCR fragments spanning exons 5- 8 may be amplified from clinical samples using the technique of RT-PCR (reverse transcription-PCR); kits which permit the user to start with tissue and produce a PCR product are available from Perkin Elmer (Norwalk, CT) and Stratagene (LaJolla, CA). The RT-PCR technique generates a single-stranded cDNA corresponding to a chosen segment of the coding region of a gene by using reverse transcription of RNA; the single-stranded cDNA is then used as template in the PCR. In the case of the p53 gene, an approximately 600 bp fragment spanning exons 5-8 is generated using primers located in the coding region immediately adjacent to exons 5 and 8 in the RT-PCR. The PCR amplified segment is then subjected to the CFLP reaction and the reaction products are analysed as described above in section VIII.

Fragments suitable for CFLP analysis may also be generated by PCR amplification of genomic DNA. DNA is extracted from a sample and primers corresponding to sequences present in introns 4 and 8 are used to amplify a segment of the p53 gene spanning exons 5-8 which includes introns 5-7 (an approximately 2 kb fragment). If it is desirable to use smaller fragments of DNA in the CFLP reaction,

primers may be chosen to amplify smaller (1 kb or less) segments of genomic DNA or alternatively a large PCR fragment may be divided into two or more smaller frgaments using resstriction enzymes.

In order to facilitate the identification of p53 mutations in the clinical setting, a library containing the CFLP pattern produced by previously characterized mutations may be provided. Comparision of the pattern generated using nucleic acid derived from a clinical sample with the patterns produced by cleavage of known and characterized p53 mutations will allow the rapid identification of the specific p53 mutation present in the patient's tissue. The comparison of CFLP patterns from clinical samples to the patterns present in the library may be accomplished by a variety of means. The simplest and least expensive comparision involves visual comparision. Given the large number of unique mutations known at the p53 locus, visual (*i.e.*, manual) comparision may be too time-consuming, especially when large numbers of clinical isolates are to be screened. Therefore the CFLP patterns or "bar codes" may be provided in an electronic format for ease and efficiency in comparision. Electronic entry may comprise storage of scans of gels containing the CFLP products of the reference p53 mutations (using for example, the GeneReader and Gel Doctor Fluorescence Gel documentation system (BioRad, Hercules, CA) or the ImageMaster (Pharmacia Biotech, Piscataway, NJ). Alternatively, as the detection of cleavage patterns may be automated using DNA sequencing instrumentation (see Example 20), the banding pattern may be stored as the signal collected from the appropriate channels during an automated run [examples of instrumentation suitable for such analysis and data collection include fluorescence-based gel imagers such as fluoroimagers produced by Molecular Dynamics and Hitachi or by real-time electrophoresis detection systems such as the ABI 377 or Pharmacia ALF DNA Sequencer].

## B. Generation of a Library of Characterized p53 Mutations

The generation of a library of characterized mutations will enable clinical samples to be rapidly and directly screened for the presence of the most common p53 mutations. Comparision of CFLP patterns generated from clinical samples to the p53 bar code library will establish both the presence of a mutation in the p53 gene and its precise identity without the necessity of costly and time consuming DNA sequence analysis.

The p53 bar code library is generated using reverse genetics. Engineering of p53 mutations ensures the identity and purity of each of the mutations as each engineered mutation is confirmed by DNA sequencing. The individual p53 mutations in p53 bar code library are generated using the 2-step "recombinant PCR" technique [Higuchi, R. (1991) In Ehrlich, H.A. (Ed.), PCR Technology: Principles and Applications for DNA Amplification, Stockton Press, New York, pp. 61-70 and Nelson, R.M. and Long, G.L. (1989) Analytical Biochem. 180:147]. Figure 77 provides a schematic representation of one method of a 2-step recombinant PCR technique that may be used for the generation of p53 mutations.

The template for the PCR amplifications is the entire human p53 cDNA gene. In the first of the two PCRs (designated "PCR 1" in Fig. 77), an oligonucleotide containing the engineered mutation ("oligo A" in Fig. 77) and an oligonucleotide containing a 5' arm of approximately 20 non-complementary bases ("oligo B") are used to amplify a relatively small region of the target DNA (100-200 bp). The resulting amplification product will contain the mutation at its extreme 5' end and a foreign sequence at its 3' end. The 3' sequence is designed to include a unique restriction site (*e.g.*, Eco RI) to aid in the directional cloning of the final amplification fragment (important for purposes of sequencing and archiving the DNA containing the mutation). The product generated in the upstream or first PCR may be gel purified if desired prior to the use of this first PCR product in the second PCR; however gel purification is not required once it is established that this fragment is the only species amplified in the PCR.

The small PCR fragment containing the engineered mutation is then used to direct a second round of PCR (PCR 2). In PCR 2, the target DNA is a larger fragment (approximately 1 kb) of the same subcloned region of the p53 cDNA. Because the sequence at the 3' end of the small PCR fragment is not complementary to any of the sequences present in the target DNA, only that strand in which the mismatch is at the extreme 5' end is amplified in PCR 2 (a 3' non-templated arm cannot be extended in PCR). Amplification is accomplished by the addition of a primer complementary to a region of the target DNA upstream of the locus of the engineered mutation ("oligo C") and by the addition of a primer complementary to the 5' noncomplementary sequence of the small product of PCR 1 ("oligo D"). By directing amplification from the noncomplementary sequence, this procedure results in the specific amplification of only those sequences containing the mutation. In order to facilitate cloning of these PCR products into a standard vector, a second unique restriction site can be engineered into oligo C (*e.g., Hind*III).

The use of this 2-step PCR approach requires that only one primer be synthesized for each mutant to be generated after the initial set-up of the system (*i.e.*, oligo A). Oligos B, C and D can be used for all mutations generated within a given region. Because oligos C and D are designed to include different and unique restriction sites, subsequent directional cloning of these PCR products into plasmid vectors (such as pUC 19) is greatly simplified. Selective amplification of only those sequences that include the desired mutational change simplifies identification of mutation-containing clones as only verification of the sequence of insert containing plasmids is required. Once the sequence of the insert has been verified, each mutation-containing clone may be maintained indefinitely as a bacterial master stock. In addition, DNA stocks of each mutant can be maintained in the form of large scale PCR preparations. This permits distribution of either bactera harboring plasmids containing a given mutation or a PCR preparation to be distributed as individual controls in kits containing reagents for the scanning of p53 mutations in clinical samples or as part of a supplemental master p53 mutation library control kit.

An alternative 2-step recombinant PCR is diagrammed in Figure 78, and described in Example 32. In this method two mutagenic oligonucleotides, one for each strand, are synthesized. These oligonucleotides are substantially complementary to each other but are opposite in orientation.. That is, one is positioned to allow amplification of an "upstream" region of the DNA, with the mutation incorporated into the 3' proximal region of the upper, or sense strand, while the other is positioned to allow amplification of a "downstream" segment with the intended mutation incorporated into the 5' proximal region of the upper, or sense strand. These two double stranded products share the sequence provided by these mutagenic oligonucleotides. When purified, combined, denatured and annealed, those strands that anneal with recessed 3' ends can be extended or filled in by the action of DNA polymerase, thus recreating a full length molecules with the mutation in the central region. This recombinant can be amplified by the use of the "outer" primer pair,those used to make the 5' end of the "upstream" and the 3' end of the "downstream" intermediate fragments.

While extra care must be taken with this method (in comparison with the method described above) because the outer primers can amplify both the recombinant and the un-modified sequence, this method does allow rapid recombinant PCR to be performed using existing end primers, and without the introduction of foreign sequences. In summary, this method is often used if only a few recombinations are to be performed. When large volumes of mutagenic PCRs are to be performed, the first described method is preferable as the first method requires a single oligo be synthesized for each mutagenesis and only recombinants are amplified.

An important feature of kits designed for the identifcation of p53 mutations in clinical samples is the inclusion of the specific primers to be used for generating PCR fragments to be analyzed for CFLP. While DNA fragments from 100 to over 1500 bp can be reproducibly and accurately analyzed for the presence of sequence polymorphisms by this technique, the precise patterns generated from different length fragments of the same input DNA sequence will of course vary. Not only are patterns shifted relative to one another depending on the length of the input DNA, but in some

cases, more long range interactions between distant regions of long DNA fragments may result in the generation of additional cleavage products not seen with shorter input DNA products. For this reason, exact matches with the bar code library will be assured through the use of primers designed to amplify the same size fragment from the clinical samples as was used to generate a given version of the p53 bar code library.

### C. Detection of Unique CFLP™ Patterns for p53 Mutations

The simplest and most direct method of analyzing the DNA fragments produced in the CFLP™ reaction is by gel electrophoresis. Because electrophoresis is widely practiced and easily accessible, initial efforts have been aimed at generating a database in this familiar format. It should, however, be noted that resolution of DNA fragments generated by CFLP™ analysis is not limited to electrophoretic methods. Mass spectrometry, chromotography, fluorescence polarization, and chip hybridization are all approaches that are currently being refined and developed in a number of research laboratories. Once generated, the CFLP™ database is easily adapted to analysis by any of these methods.

There are several possible alternatives available for detection of CFLP patterns. A critical user benefit of CFLP analysis is that the results are not dependent on the chosen method of DNA detection. DNA fragments may be labeled with a radioisotope (*e.g.*, a $^{32}P$ or $^{35}S$-labeled nucleotide) placed at either the 5' or 3' end of the nucleic acid or alternatively the label may be distributed throughout the nucleic acid *(i.e.*, an internally labeled substrate). The label may be a nonisotopic detectable moiety, such as a fluorophore which can be detected directly, or a reactive group which permits specific recognition by a secondary agent. CFLP patterns have been detected by immunostaining, biotin-avidin interactions, autoradiography and direct fluorescence imaging. Since radiation use is in rapid decline in clinical settings and since both immunostaining and biotin-avidin based detection schemes require time-consuming

transfer of DNA onto an expensive membrane support, fluorescence-based detection methods may be preferred. It is important to note, however, that any of the above methods may be used to generate CFLP bar codes to be input into the database.

In addition to their being a direct, non-isotopic means of detecting CFLP patterns, fluorescence-based schemes offer a noteworthy additional advantage in clinical applications. CFLP allows the analysis of several samples in the same tube and in the same lane on a gel. This "multiplexing" permits rapid and automated comparison of a large number of samples in a fraction of the time and for a lower cost than can be realized through individual analysis of each sample. This approach opens the door to several alternative applications. A researcher could decide to double, triple or quadruple (up to 4 dyes have been demonstrated to be detectable and compatible in a single lane in commerically available DNA sequencing instrumentation such as the ABI 373/377) the number of samples run on a given gel. Alternatively, the analyst may include a normal p53 gene sample in each tube, and each gel lane, along with a differentially labeled size standard, as a internal standard to verify both the presence and the exact location(s) of a pattern difference(s) between the normal p53 gene and putative mutants.

## VI. Detection and Identification of Pathogens Using the CFLP™ Method

### A. Detection and Identification of Hepatitis C Virus

Hepatitis C virus (HCV) infection is the predominant cause of post-transfusion non-A, non-B (NANB) hepatitis around the world. In addition, HCV is the major etiologic agent of hepatocellular carcinoma (HCC) and chronic liver disease world wide. HCV infection is transmitted primarily to blood transfusion recipients and intravenous drug users although maternal transmission to offspring and transmission to recipients of organ transplants have been reported.

The genome of the positive-stranded RNA hepatitis C virus comprises several regions including 5' and 3' noncoding regions (*i.e.*, 5' and 3' untranslated regions) and a polyprotein coding region which encodes the core protein (C), two envelope glycoproteins (E1 and E2/NS1) and six nonstructural glycoproteins (NS2-NS5b).

Molecular biological analysis of the small (9.4 kb) RNA genome has showed that some regions of the genome are very highly conserved between isolates, while other regions are fairly rapidly changeable. The 5' noncoding region (NCR) is the most highly conserved region in the HCV. These analyses have allowed these viruses to be divided into six basic genotype groups, and then further classified into over a dozen sub-types [the nomenclature and division of HCV genotypes is evolving; see Altamirano *et al., J. Infect. Dis.* 171:1034 (1995) for a recent classification scheme]. These viral groups are associated with different geographical areas, and accurate identification of the agent in outbreaks is important in monitoring the disease. While only Group 1 HCV has been observed in the United States, multiple HCV genotypes have been observed in both Europe and Japan.

The ability to determine the genotype of viral isolates also allows comparisons of the clinical outcomes from infection by the different types of HCV, and from infection by multiple types in a single individual. HCV type has also been associated with differential efficacy of treatment with interferon, with Group 1 infected individuals showing little response [Kanai *et al., Lancet* 339:1543 (1992) and Yoshioka *et al., Hepatology* 16:293 (1992)]. Pre-screening of infected individuals for the viral type will allow the clinician to make a more accurate diagnosis, and to avoid costly but fruitless drug treatment.

Existing methods for determining the genotype of HCV isolates include PCR amplification of segments of the HCV genome coupled with either DNA sequencing or hybridization to HCV-specific probes, RFLP analysis of PCR amplified HCV DNA anything else?. All of these methods suffer from the limitations discussed above (*i.e.*, DNA sequencing is too labor-intensive and expensive to be practical in clinical laboratory settings; RFLP analysis suffers from low sensitivity).

Universal and genotype specific primers have been designed for the amplification of HCV sequences from RNA extracted from plasma or serum [Okamoto *et al. J. Gen. Virol.* 73:673 (1992);Yoshioka *et al., Hepatology* 16:293 (1992) and Altamirano *et al., supra*]. These primers can be used to generate PCR products which

serve as substrates in the CFLP™ assay of the present invention. As shown herein CFLP™ analysis provides a rapid and accurate method of typing HCV isolates. CFLP™ analysis of HCV substrates allows a distinction to be made between the major genotypes and subtypes of HCV thus providing improved methods for the genotyping of HCV isolates.

**B. Detection and Identification of Multi-Drug Resistant *M. tuberculosis***

In the past decade there has been a tremendous resurgence in the incidence of tuberculosis in this country and throughout the world. In the United States, the incidence of tuberculosis has risen steadily during past decade, accounting for 2000 deaths annually, with as many as 10 million Americans infected with the disease. The situation is critical in New York City, where the incidence has more than doubled in the past decade, accounting for 14% of all new cases in the United States in 1990 [Frieden *et al., New Engl. J. Med.* 328:521 (1993)].

The crisis in New York City is particularly dire because a significant proportion (as many as one-third) of the recent cases are resistant to one or more antituberculosis drugs [Frieden *et al, supra* and Hughes, *Scrip Magazine* May (1994)]. Multi-drug resistant tuberculosis (MDR-TB) is an iatrogenic disease that arises from incomplete treatment of a primary infection [Jacobs, Jr., *Clin. Infect. Dis.* 19:1 (1994)]. MDR-TB appears to pose an especially serious risk to the immunocompromised, who are more likely to be infected with MDR-TB strains than are otherwise healthy individuals [Jacobs, Jr., *supra*]. The mortality rate of MDR-TB in immunocompromised individuals is alarmingly high, often exceeding 90%, compared to a mortality rate of <50% in otherwise uncompromised individuals [Donnabella *et al., Am. J. Respir. Dis.* 11:639 (1994)].

From a clinical standpoint, tuberculosis has always been difficult to diagnose because of the extremely long generation time of *Mycobacterium tuberculosis* as well as the environmental prevalence of other, faster growing mycobacterial species. The doubling time of *M. tuberculosis* is 20-24 hours, and growth by conventional methods

typically requires 4 to 6 weeks to positively identify *M. tuberculosis* [Jacobs, Jr. *et al.*, *Science* 260:819 (1993) and Shinnick and Jones in *Tuberculosis: Pathogenesis, Protection and Control*, Bloom, ed., American Society of Microbiology, Washington, D.C. (1994), pp. 517-530]. It can take an additional 3 to 6 weeks to diagnose the drug susceptibility of a given strain [Shinnick and Jones, *supra*]. Needless to say, the health risks to the infected individual, as well as to the public, during a protracted period in which the patient may or may not be symptomatic, but is almost certainly contagious, are considerable. Once a drug resistance profile has been elucidated and a diagnosis made, treatment of a single patient can cost up to $250,000 and require 24 months.

The recent explosion int he incidence of the disease, together with the dire risks posed by MDR strains, have combined to spur a burst of research activity and commercial development of procedures and products aimed at accelerating the detection of *M. tuberculosis* as well the elucidation of drug resistance profiles of *M. tuberculosis* clinical isolates. A number of these methods are devoted primarily to the task of determining whether a given strain is *M. tuberculosis* or a mycobacterial species other than tuberculosis. Both culture based methods and nucleic-acid based methods have been developed that allow *M. tuberculosis* to be positively identified more rapidly than by classical methods: detection times have been reduced from greater than 6 weeks to as little as two weeks (culture-based methods) or two days (nucleic acid-based methods). While culture-based methods are currently in wide-spread use in clinical laboratories, a number of rapid nucleic acid-based methods that can be applied directly to clinical samples are under development. For all of the techniques described below, it is necessary to first "decontaminate" the clinical samples, such as sputum (usually done by pretreatment with N-acetyl L-cysteine and NaOH) to reduce contamination by non-mycobacterial species [Shinnick and Jones, *supra.*]

The polymerase chain reaction (PCR) has been applied to the detection of *M. tuberculosis* and can be used to detect its presence directly from clinical specimens within one to two days. The more sensitive techniques rely on a two-step procedure:

the first step is the PCR amplification itself, the second is an analytical step such as hybridization of the amplicon to a *M. tuberculosis*-specific oligonucleotide probe, or analysis by RFLP or DNA sequencing [Shinnick and Jones, *supra*].

The Amplified *M. tuberculosis* Direct Test (AMTDT; Gen-Probe) relies on Transcription Mediated Amplification [TMA; essentially a self-sustained sequence reaction (3SR) amplification] to amplify target rRNA sequences directly from clinical specimens. Once the rRNA has been amplified, it is then detected by a dye-labeled assay such as the PACE2. This assay is highly subject to inhibition by substances present in clinical samples.

The Cycling Probe Reaction (CPR; ID Biomedical). This technique, which is under development as a diagnostic tool for detecting the presence of *M. tuberculosis*, measures the accumulation of signal probe molecules. The signal amplification is accomplished by hybridizing tripartite DNA-RNA-DNA probes to target nucleic acids, such as *M. tuberculosis*-specific sequences. Upon the addition of RNAse H, the RNA portion of the chimeric probe is degraded, releasing the DNA portions, which accumulate linearly over time to indicate that the target sequence is present [Yule, Bio/Technology 12:1335 (1994)]. The need to use of RNA probes is a drawback, particularly for use in crude clinical samples, where RNase contamination is often rampant.

The above nucleic acid-based detection and differentiation methods offer a clear time savings over the more traditional, culture-based methods. While they are beginning to enter the clinical setting, their usefulness in the routine diagnosis of *M. tuberculosis* is still in question, in large part because of problems with associated with cross-contamination and low-sensitivity relative to culture-based methods. In addition, many of these procedures are limited to analysis of respiratory specimens [Yule, Bio/Technology 12:1335 (1994)].

ii) Determination of the antibiotic resistance profile of *M. tuberculosis*

a) Culture-based methods: Once a positive identification of *M. tuberculosis* has been made, it is necessary to characterize the extent and nature of the

strain's resistance to antibiotics. The traditional method used to determine antibiotic resistance is the direct proportion agar dilution method, in which dilutions of culture are plated on media containing antibiotics and on control media without antibiotics. This method typically adds an additional 2-6 weeks to the time required for diagnosis and characterization of an unknown clinical sample [Jacobs, Jr., *supra*].

The Luciferase Reporter Mycobacteriophage (LRM) assay was first described in 1993 [Jacobs, Jr. *et al., Science* 260:819 (1993)]. In this assay, a mycobacteriophage containing a cloned copy of the luciferase gene is used to infect mycobacterial cultures. In the presence of luciferin and ATP, the expressed luciferase produces photons, easily distinguishable by eye or by a luminometer, allowing a precise determination of the extent of mycobacterial growth in the presence of antibiotics. Once sufficient culture has been obtained (usually 10-14 days post-inoculation), the assay can be completed in 2 days. This method suffers from the fact that the LRM are not specific for *M. tuberculosis*: they also infect *M. smegmatis* and *M. bovis* (*e.g.*, BCG), thereby complicating the interpretation of positive results. Discrimination between the two species must be accomplished by growth on specialized media which does not support the growth of *M. tuberculosis* (*e.g.*, NAP media). This confirmation requires another 2 to 4 days.

The above culture-based methods for determining antibiotic resistance will continue to play a role in assessing the effectiveness of putative new anti-mycobacterial agents and those drugs for which a genetic target has not yet been identified. However, recent success in elucidating the molecular basis for resistance to a number of anti-mycobacterial agents, including many of the front-line drugs, has made possible the use of much faster, more accurate and more informative DNA polymorphism-based assays.

b) DNA-based methods: Genetic loci involved in resistance to isoniazid, rifampin, streptomycin, fluoroquinolones, and ethionamide have been identified [Jacobs, Jr., *supra*; Heym *et al., Lancet* 344:293 (1994) and Morris *et al., J. Infect. Dis.* 171:954 (1995)]. A combination of isoniazid (inh) and rifampin (rif) along

with pyrazinamide and ethambutol or streptomycin, is routinely used as the first line of attack against confirmed cases of *M. tuberculosis* [Banerjee *et al., Science* 263:227 (1994)]. Consequently, resistance to one or more of these drugs can have disastrous implications for short course chemotherapy treatment. The increasing incidence of such resistant strains necessitates the development of rapid assays to detect them and thereby reduce the expense and community health hazards of pursuing ineffective, and possibly detrimental, treatments. The identification of some of the genetic loci involved in drug resistance has facilitated the adoption of mutation detection technologies for rapid screening of nucleotide changes that result in drug resistance. The availability of amplification procedures such as PCR and SDA, which have been successful in replicating large amounts of target DNA directly from clinical specimens, makes DNA-based approaches to antibiotic profiling far more rapid than conventional, culture-based methods.

The most widely employed techniques in the genetic identification of mutations leading to drug resistance are DNA sequencing, Restriction Fragment Length Polymorphism (RFLP), PCR-Single Stranded Conformational Polymorphism (PCR-SSCP), and PCR-dideoxyfingerprinting (PCR-ddF). All of these techniques have drawbacks as discussed above. None of them offers a rapid, reproducible means of precisely and uniquely identifying individual alleles.

In contrast the CFLP™ method of the present invetion provides an approach that relies on structure specific cleavage to generate distinct collections of DNA fragments. This method is highly sensitive (>98%) in its ability to detect sequence polymorphisms, and requires a fraction of the time, skill and expense of the techniques described above.

The application of the CFLP™ method to the detection of MDR-TB is illustrated herein using segments of DNA amplified from the *rpoB* and *katG* genes. Other genes associated with MDR-TB, including but not limited to those involved in conferring resistance to isoniazid (*inhA*), streptomycin (*rpsL* and *rrs*), and fluoroquinoline (*gyrA*), are equally well suited to the CFLP™ assay.

### C. Detection and Identification of Bacterial Pathogens

Identification and typing of bacterial pathogens is critical in the clinical management of infectious diseases. Precise identity of a microbe is used not only to differentiate a disease state from a healthy state, but is also fundamental to determining whether and which antibiotics or other antimicrobial therapies are most suitable for treatment. Traditional methods of pathogen typing have used a variety of phenotypic features, including growth characteristics, color, cell or colony morphology, antibiotic susceptibility, staining, smell and reactivity with specific antibodies to identify bacteria. All of these methods require culture of the suspected pathogen, which suffers from a number of serious shortcomings, including high material and labor costs, danger of worker exposure, false positives due to mishandling and false negatives due to low numbers of viable cells or due to the fastidious culture requirements of many pathogens. In addition, culture methods require a relatively long time to achieve diagnosis, and because of the potentially life-threatening nature of such infections, antimicrobial therapy is often started before the results can be obtained. In many cases the pathogens are very similar to the organisms that make up the normal flora, and may be indistinguishable from the innocuous strains by the methods cited above. In these cases, determinion of the presence of the pathogenic strain may require the higher resolution afforded by more recently developed molecular typing methods.

A number of methods of examining the genetic material from organisms of interest have been developed. One way of performing this type of analysis is by hybridization of species-specific nucleic acid probes to the DNA or RNA from the organism to be tested. This may be done by immobilizing the denatured nucleic acid

to be tested on a membrane support, and probing with labeled nucleic acids that will bind only in the presence of the DNA or RNA from the pathogen. In this way, pathogens can be identified. Organisms can be further diffrentiated by using the RFLP method described above, in which the genomic DNA is digested with one or more restriction enzymes before electrophoretic separation and transfer to a nitrocellulose or nylon membrane support. Probing with the species-specific nucleic acid probes will reveal a banding pattern that, if it shows variation between isolates, can be used as a reproducible way of discriminating between strains. However, these methods are susceptible to the drawbacks outlined above: hybridization-based assays are time-consuming and may give false or misleading results if the stringency of the hybridization is not well controlled, and RFLP identification is dependent on the presence of suitable restriction sites in the DNA to be analyzed.

To address these concerns about hybridization and RFLP as diagnostic tools, several methods of molecular analysis based on polymerase chain reaction (PCR) amplification have gained popularity. In one well-accepted method, called PCR fingerprinting, the size of a fragment generated by PCR is used as an identifier. In this type of assay, the primers are targeted to regions containing variable numbers of tandem repeated sequences (referred to as VNTRs an eukaryotes). The number of repeats, and thus the length of the PCR amplicon, can be characteristic of a given pathogen, and co-amplification of several of these loci in a single reaction can create specific and reproducible fingerprints, allowing discrimination between closely related species.

In some cases where organisms are very closely related, however, the target of the amplification does not display a size difference, and the amplified segment must be further probed to achieve more precise identification. This may be done on a solid support, in a fashion analogous to the whole-genome hybridization described above, but this has the same problem with variable stringency as that assay. Alternatively, the interior of the PCR fragment may be used as a template for a sequence-specific ligation event. As outlined above for the LCR, in this method, single stranded probes to be ligated are positioned along the sequence of interest on either side of an

identifying polymorphism, so that the success or failure of the ligation will indicate the presence or absence of a specifice nucleotide sequence at that site. With either hybridization or ligation methods of PCR product analysis, knowledge of the precise sequence in the area of probe binding must be obtained in advance, and differences outside the probe binding area are not detected. These methods are poorly suited to the examination and typing of new isolates that have not been fully characterized.

In the methods of the present invention, primers that recognize conserved regions of bacterial ribosomal RNA genes allow amplification of segments of these genes that include sites of variation. The variations in ribosomal gene sequences have become an accepted method not only of differenting between similar organisms on a DNA sequence level, but their consistant rate of change allows these sequences to be used to evaluate the evolutionary relatedness of orgnaisms. That is to say, the more similar the nucleic acid is at the sequence level, the more closely related the organisms in discussion are considered to be. [Woese, Bacterial Evolution. Microbiological Reviews, vol 51, No. 2. 1987]. The present invention allows the amplification products derived from these sequences to be used to create highly individual barcodes (*i.e.*, cleavage patterns), allowing the detection of sequence polymorphisms without prior knowledge of the site, character or even the presence of said polymorphisms. With appropriate selection of primers, amplification can be made to be either all-inclusive (*e.g.*, using the most highly conserved ribosomal sequences) to allow comparison of distantly related organisms, or the primers can be chosen to be very specific for a given genus, to allow examination at the species and subspecies level. While the examination of ribosomal genes is extremely useful in these characterizations, the use of the CFLP™ method in bacterial typing is not limited to these genes. Other genes, including but not limited to those associated with specifc growth characterisics, (*e.g.*, carbon source preference, antibiotic resistance, resistance to methycillin or antigen production), or with particular cell morphologies (such as pilus formation) are equally well suited to the CFLP™ assay.

## D. Extraction of Nucleic Acids From Clinical Samples

To provide nucleic acid substrates for use in the detection and identification of microorganisms in clinical samples using the CFLP™ assay, nucleic acid is extracted from the sample. The nucleic acid may be extracted from a variety of clinical samples [fresh or frozen tissue, suspensions of cells (*e.g.*, blood), cerebral spinal fluid, sputum, urine, etc.] using a variety of standard techniques or commercially available kits. For example, kits which allow the isolation of RNA or DNA from tissue samples are available from Qiagen, Inc. (Chatsworth, CA) and Stratagene (LaJolla, CA). For example, the QIAamp Blood kits permit the isolation of DNA from blood (fresh, frozen or dried) as well as bone marrow, body fluids or cell suspensions. QIAamp tissue kits permit the isolation of DNA from tissues such as muscles, organs and tumors.

It has been found that crude extracts from relatively homogenous specimens (such as blood, bacterial colonies, viral plaques, or cerebral spinal fluid) are better suited to severing as templates for the amplification of unique PCR products than are more composite specimens (such as urine, sputum or feces;) [Shibata in *PCR:The Polymerase Chain Reaction*, Mullis et al., eds., Birkhauser, Boston (1994), pp. 47-54]. Samples which contain relatively few copies of the material to be amplified (*i.e.*, the target nucleic acid), such as cerebral spinal fluid, can be added directly to a PCR. Blood samples have posed a special problem in PCRs due to the inhibitory properties of red blood cells. The red blood cells must be removed prior to the use of blood in a PCR; there are both classical and commercially available methods for this purpose [*e.g.*, QIAamp Blood kits, passage through a Chelex 100 column (BioRad), etc.]. Extraction of nucleic acid from sputum, the specimen of choice for the direct detection of M. tuberculosis, requires prior decontamination to kill or inhibit the growth of other bacterial species. This decontamination is typically accomplished by treatment of the sample with N-acetyl L-cysteine and NaOH (Shinnick and Jones, *supra*). This decontamination process is necessary only when the sputum specimen is to be cultured prior to analysis.

## EXPERIMENTAL

The following examples serve to illustrate certain preferred embodiments and aspects of the present invention and are not to be construed as limiting the scope thereof.

In the disclosure which follows, the following abbreviations apply:°C (degrees Centigrade); g (gravitational field); vol (volume); w/v (weight to volume); v/v (volume to volume); BSA (bovine serum albumin); CTAB (cetyltrimethylammonium bromide); HPLC (high pressure liquid chromatography); DNA (deoxyribonucleic acid); IVS (intervening sequence); p (plasmid); µl (microliters); ml (milliliters); µg (micrograms); pmoles (picomoles); mg (milligrams); MOPS (3-[N-Morpholino]propanesulfonic acid); M (molar); mM (milliMolar); µM (microMolar); nm (nanometers); kdal (kilodaltons); OD (optical density); EDTA (ethylene diamine tetra-acetic acid); FITC (fluorescein isothiocyanate); SDS (sodium dodecyl sulfate); $NaPO_4$ (sodium phosphate); Tris (tris(hydroxymethyl)-aminomethane); PMSF (phenylmethylsulfonylfluoride); TBE (Tris-Borate-EDTA, *i.e.*, Tris buffer titrated with boric acid rather than HCl and containing EDTA) ; PBS (phosphate buffered saline); PPBS (phosphate buffered saline containing 1 mM PMSF); PAGE (polyacrylamide gel electrophoresis); Tween (polyoxyethylene-sorbitan); Boehringer Mannheim (Boehringer Mannheim, Indianapolis, IN); Dynal (Dynal A.S., Oslo, Norway); Epicentre (Epicentre Technologies, Madison, WI); National Biosciences (National Biosciences, Plymouth, MN); New England Biolabs (New England Biolabs, Beverly, MA); Novagen (Novagen, Inc., Madison, WI); Perkin Elmer (Perkin Elmer, Norwalk, CT); Promega Corp. (Promega Corp., Madison, WI); RJ Research (RJ Research, Inc., Watertown, MA); Stratagene (Stratagene Cloning Systems, La Jolla, CA); USB (U.S. Biochemical, Cleveland, OH).

## EXAMPLE 1

Characteristics Of Native Thermostable DNA Polymerases

### A. 5' Nuclease Activity Of DNAP*Taq*

During the polymerase chain reaction (PCR) [Saiki *et al.*, *Science* 239:487 (1988); Mullis and Faloona, *Methods in Enzymology* 155:335 (1987)], DNAP*Taq* is able to amplify many, but not all, DNA sequences. One sequence that cannot be amplified using DNAP*Taq* is shown in Figure 6 (Hairpin structure is SEQ ID NO:15, PRIMERS are SEQ ID NOS:16-17.) This DNA sequence has the distinguishing characteristic of being able to fold on itself to form a hairpin with two single-stranded arms, which correspond to the primers used in PCR.

To test whether this failure to amplify is due to the 5' nuclease activity of the enzyme, we compared the abilities of DNAP*Taq* and DNAPStf to amplify this DNA sequence during 30 cycles of PCR. Synthetic oligonucleotides were obtained from The Biotechnology Center at the University of Wisconsin-Madison. The DNAP*Taq* and DNAPStf were from Perkin Elmer (*i.e.*, AmpliTaq™ DNA polymerase and the Stoffel fragment of Amplitaq™ DNA polymerase). The substrate DNA comprised the hairpin structure shown in Figure 6 cloned in a double-stranded form into pUC19. The primers used in the amplification are listed as SEQ ID NOS:16-17. Primer SEQ ID NO:17 is shown annealed to the 3' arm of the hairpin structure in Fig. 6. Primer SEQ ID NO:16 is shown as the first 20 nucleotides in bold on the 5' arm of the hairpin in Fig. 6.

Polymerase chain reactions comprised 1 ng of supercoiled plasmid target DNA, 5 pmoles of each primer, 40 μM each dNTP, and 2.5 units of DNAP*Taq* or DNAPStf, in a 50 μl solution of 10 mM Tris•Cl pH 8.3. The DNAP*Taq* reactions included 50 mM KCl and 1.5 mM $MgCl_2$. The temperature profile was 95°C for 30 sec., 55°C for 1 min. and 72°C for 1 min., through 30 cycles. Ten percent of each reaction was analyzed by gel electrophoresis through 6% polyacrylamide (cross-linked 29:1) in a buffer of 45 mM Tris•Borate, pH 8.3, 1.4 mM EDTA.

The results are shown in Figure 7. The expected product was made by DNAPStf (indicated simply as "S") but not by DNAP*Taq* (indicated as "T"). We conclude that the 5' nuclease activity of DNAP*Taq* is responsible for the lack of amplification of this DNA sequence.

To test whether the 5' unpaired nucleotides in the substrate region of this structured DNA are removed by DNAP*Taq*, the fate of the end-labeled 5' arm during four cycles of PCR was compared using the same two polymerases (Figure. 8). The hairpin templates, such as the one described in Figure 6, were made using DNAPStf and a $^{32}$P-5'-end-labeled primer. The 5'-end of the DNA was released as a few large fragments by DNAP*Taq* but not by DNAPStf. The sizes of these fragments (based on their mobilities) show that they contain most or all of the unpaired 5' arm of the DNA. Thus, cleavage occurs at or near the base of the bifurcated duplex. These released fragments terminate with 3' OH groups, as evidenced by direct sequence analysis, and the abilities of the fragments to be extended by terminal deoxynucleotidyl transferase.

Figures 9-11 show the results of experiments designed to characterize the cleavage reaction catalyzed by DNAP*Taq*. Unless otherwise specified, the cleavage reactions comprised 0.01 pmoles of heat-denatured, end-labeled hairpin DNA (with the unlabeled complementary strand also present), 1 pmole primer (complementary to the 3' arm) and 0.5 units of DNAP*Taq* (estimated to be 0.026 pmoles) in a total volume of 10µl of 10 mM Tris-Cl, ph 8.5, 50 mM KCl and 1.5 mM $MgCl_2$. As indicated, some reactions had different concentrations of KCl, and the precise times and temperatures used in each experiment are indicated in the individual figures. The reactions that included a primer used the one shown in Figure 6 (SEQ ID NO:17). In some instances, the primer was extended to the junction site by providing polymerase and selected nucleotides.

Reactions were initiated at the final reaction temperature by the addition of either the $MgCl_2$ or enzyme. Reactions were stopped at their incubation temperatures by the addition of 8 µl of 95% formamide with 20 mM EDTA and 0.05% marker

dyes. The $T_m$ calculations listed were made using the Oligo™ primer analysis software from National Biosciences, Inc. These were determined using 0.25 μM as the DNA concentration, at either 15 or 65 mM total salt (the 1.5 mM $MgCl_2$ in all reactions was given the value of 15 mM salt for these calculations).

Figure 9 is an autoradiogram containing the results of a set of experiments and conditions on the cleavage site. Figure 9A is a determination of reaction components that enable cleavage. Incubation of 5'-end-labeled hairpin DNA was for 30 minutes at 55°C, with the indicated components. The products were resolved by denaturing polyacrylamide gel electrophoresis and the lengths of the products, in nucleotides, are indicated. Figure 9B describes the effect of temperature on the site of cleavage in the absence of added primer. Reactions were incubated in the absence of KCl for 10 minutes at the indicated temperatures. The lengths of the products, in nucleotides, are indicated.

Surprisingly, cleavage by DNAP*Taq* requires neither a primer nor dNTPs (*see* Fig. 9A). Thus, the 5' nuclease activity can be uncoupled from polymerization. Nuclease activity requires magnesium ions, though manganese ions can be substituted, albeit with potential changes in specificity and activity. Neither zinc nor calcium ions support the cleavage reaction. The reaction occurs over a broad temperature range, from 25°C to 85°C, with the rate of cleavage increasing at higher temperatures.

Still referring to Figure 9, the primer is not elongated in the absence of added dNTPs. However, the primer influences both the site and the rate of cleavage of the hairpin. The change in the site of cleavage (Fig. 9A) apparently results from disruption of a short duplex formed between the arms of the DNA substrate. In the absence of primer, the sequences indicated by underlining in Figure 6 could pair, forming an extended duplex. Cleavage at the end of the extended duplex would release the 11 nucleotide fragment seen on the Fig. 9A lanes with no added primer. Addition of excess primer (Fig. 9A, lanes 3 and 4) or incubation at an elevated temperature (Fig. 9B) disrupts the short extension of the duplex and results in a longer 5' arm and, hence, longer cleavage products.

The location of the 3' end of the primer can influence the precise site of cleavage. Electrophoretic analysis revealed that in the absence of primer (Fig. 9B), cleavage occurs at the end of the substrate duplex (either the extended or shortened form, depending on the temperature) between the first and second base pairs. When the primer extends up to the base of the duplex, cleavage also occurs one nucleotide into the duplex. However, when a gap of four or six nucleotides exists between the 3' end of the primer and the substrate duplex, the cleavage site is shifted four to six nucleotides in the 5' direction.

Fig. 10 describes the kinetics of cleavage in the presence (Fig. 10A) or absence (Fig. 10B) of a primer oligonucleotide. The reactions were run at 55°C with either 50 mM KCl (Fig. 10A) or 20 mM KCl (Fig. 10B). The reaction products were resolved by denaturing polyacrylamide gel electrophoresis and the lengths of the products, in nucleotides, are indicated. "M", indicating a marker, is a 5' end-labeled 19-nt oligonucleotide. Under these salt conditions, Figs. 10A and 10B indicate that the reaction appears to be about twenty times faster in the presence of primer than in the absence of primer. This effect on the efficiency may be attributable to proper alignment and stabilization of the enzyme on the substrate.

The relative influence of primer on cleavage rates becomes much greater when both reactions are run in 50 mM KCl. In the presence of primer, the rate of cleavage increases with KCl concentration, up to about 50 mM. However, inhibition of this reaction in the presence of primer is apparent at 100 mM and is complete at 150 mM KCl. In contrast, in the absence of primer the rate is enhanced by concentration of KCl up to 20 mM, but it is reduced at concentrations above 30 mM. At 50 mM KCl, the reaction is almost completely inhibited. The inhibition of cleavage by KCl in the absence of primer is affected by temperature, being more pronounced at lower temperatures.

Recognition of the 5' end of the arm to be cut appears to be an important feature of substrate recognition. Substrates that lack a free 5' end, such as circular M13 DNA, cannot be cleaved under any conditions tested. Even with substrates having defined 5' arms, the rate of cleavage by DNAP*Taq* is influenced by the length

of the arm. In the presence of primer and 50 mM KCl, cleavage of a 5' extension that is 27 nucleotides long is essentially complete within 2 minutes at 55°C. In contrast, cleavages of molecules with 5' arms of 84 and 188 nucleotides are only about 90% and 40% complete after 20 minutes. Incubation at higher temperatures reduces the inhibitory effects of long extensions indicating that secondary structure in the 5' arm or a heat-labile structure in the enzyme may inhibit the reaction. A mixing experiment, run under conditions of substrate excess, shows that the molecules with long arms do not preferentially tie up the available enzyme in non-productive complexes. These results may indicate that the 5' nuclease domain gains access to the cleavage site at the end of the bifurcated duplex by moving down the 5' arm from one end to the other. Longer 5' arms would be expected to have more adventitious secondary structures (particularly when KCl concentrations are high), which would be likely to impede this movement.

Cleavage does not appear to be inhibited by long 3' arms of either the substrate strand target molecule or pilot nucleic acid, at least up to 2 kilobases. At the other extreme, 3' arms of the pilot nucleic acid as short as one nucleotide can support cleavage in a primer-independent reaction, albeit inefficiently. Fully paired oligonucleotides do not elicit cleavage of DNA templates during primer extension.

The ability of DNAP*Taq* to cleave molecules even when the complementary strand contains only one unpaired 3' nucleotide may be useful in optimizing allele-specific PCR. PCR primers that have unpaired 3' ends could act as pilot oligonucleotides to direct selective cleavage of unwanted templates during preincubation of potential template-primer complexes with DNAP*Taq* in the absence of nucleoside triphosphates.

### B. 5' Nuclease Activities Of Other DNAPs

To determine whether other 5' nucleases in other DNAPs would be suitable for the present invention, an array of enzymes, several of which were reported in the literature to be free of apparent 5' nuclease activity, were examined. The ability of

these other enzymes to cleave nucleic acids in a structure-specific manner was tested using the hairpin substrate shown in Fig. 6 under conditions reported to be optimal for synthesis by each enzyme.

DNAPEcl and DNAP Klenow were obtained from Promega Corporation; the DNAP of *Pyrococcus furious* ["Pfu", Bargseid *et al.*, Strategies 4:34 (1991)] was from Stratagene; the DNAP of *Thermococcus litoralis* ["Tli", Vent™(exo-), Perler *et al.*, Proc. Natl. Acad. Sci. USA 89:5577 (1992)] was from New England Biolabs; the DNAP of *Thermus flavus* ["Tfl", Kaledin *et al.*, *Biokhimiya* 46:1576 (1981)] was from Epicentre Technologies; and the DNAP of *Thermus thermophilus* ["Tth", Carballeira *et al.*, Biotechniques 9:276 (1990); Myers *et al.*, *Biochem.* 30:7661 (1991)] was from U.S. Biochemicals.

0.5 units of each DNA polymerase was assayed in a 20 µl reaction, using either the buffers supplied by the manufacturers for the primer-dependent reactions, or 10 mM Tris•Cl, pH 8.5, 1.5 mM $MgCl_2$, and 20mM KCl. Reaction mixtures were at held 72°C before the addition of enzyme.

Figure 11 is an autoradiogram recording the results of these tests. Figure 11A demonstrates reactions of endonucleases of DNAPs of several thermophilic bacteria. The reactions were incubated at 55°C for 10 minutes in the presence of primer or at 72°C for 30 minutes in the absence of primer, and the products were resolved by denaturing polyacrylamide gel electrophoresis. The lengths of the products, in nucleotides, are indicated. Figure 11B demonstrates endonucleolytic cleavage by the 5' nuclease of DNAPEcl. The DNAPEcl and DNAP Klenow reactions were incubated for 5 minutes at 37°C. Note the light band of cleavage products of 25 and 11 nucleotides in the DNAPEcl lanes (made in the presence and absence of primer, respectively). Figure 7B also demonstrates DNAP*Taq* reactions in the presence (+) or absence (-) of primer. These reactions were run in 50 mM and 20 mM KCl, respectively, and were incubated at 55°C for 10 minutes.

Referring to Figure 11A, DNAPs from the eubacteria *Thermus thermophilus* and *Thermus flavus* cleave the substrate at the same place as DNAP*Taq*, both in the presence and absence of primer. In contrast, DNAPs from the archaebacteria *Pyrococcus furiosus* and *Thermococcus litoralis* are unable to cleave the substrates endonucleolytically. The DNAPs from *Pyrococcus furious* and *Thermococcus litoralis* share little sequence homology with eubacterial enzymes (Ito *et al.*, *Nucl. Acids Res.* 19:4045 (1991); Mathur *et al.*, *Nucl. Acids. Res.* 19:6952 (1991); *see also* Perler *et al.*). Referring to Figure 11B, DNAPEcl also cleaves the substrate, but the resulting cleavage products are difficult to detect unless the 3' exonuclease is inhibited. The amino acid sequences of the 5' nuclease domains of DNAPEcl and DNAP*Taq* are about 38% homologous (Gelfand, *supra*).

The 5' nuclease domain of DNAP*Taq* also shares about 19% homology with the 5' exonuclease encoded by gene 6 of bacteriophage T7 [Dunn *et al.*, *J. Mol. Biol.* 166:477 (1983)]. This nuclease, which is not covalently attached to a DNAP polymerization domain, is also able to cleave DNA endonucleolytically, at a site similar or identical to the site that is cut by the 5' nucleases described above, in the absence of added primers.

### C. Transcleavage

The ability of a 5' nuclease to be directed to cleave efficiently at any specific sequence was demonstrated in the following experiment. A partially complementary oligonucleotide termed a "pilot oligonucleotide" was hybridized to sequences at the desired point of cleavage. The non-complementary part of the pilot oligonucleotide provided a structure analogous to the 3' arm of the template (*see* Figure 6), whereas the 5' region of the substrate strand became the 5' arm. A primer was provided by designing the 3' region of the pilot so that it would fold on itself creating a short hairpin with a stabilizing tetra-loop [Antao *et al.*, *Nucl. Acids Res.* 19:5901 (1991)]. Two pilot oligonucleotides are shown in Figure 12A. Oligonucleotides 19-12 (SEQ ID NO:18) and 30-12 (SEQ ID NO:19) are 31 or 42 or nucleotides long, respectively.

However, oligonucleotides 19-12 (SEQ ID NO:18) and 34-19 (SEQ ID NO:19) have only 19 and 30 nucleotides, respectively, that are complementary to different sequences in the substrate strand. The pilot oligonucleotides are calculated to melt off their complements at about 50°C (19-12) and about 75°C (30-12). Both pilots have 12 nucleotides at their 3' ends, which act as 3' arms with base-paired primers attached.

To demonstrate that cleavage could be directed by a pilot oligonucleotide, we incubated a single-stranded target DNA with DNAP*Taq* in the presence of two potential pilot oligonucleotides. The transcleavage reactions, where the target and pilot nucleic acids are not covalently linked, includes 0.01 pmoles of single end-labeled substrate DNA, 1 unit of DNAP*Taq* and 5 pmoles of pilot oligonucleotide in a volume of 20 µl of the same buffers. These components were combined during a one minute incubation at 95°C, to denature the PCR-generated double-stranded substrate DNA, and the temperatures of the reactions were then reduced to their final incubation temperatures. Oligonucleotides 30-12 and 19-12 can hybridize to regions of the substrate DNAs that are 85 and 27 nucleotides from the 5' end of the targeted strand.

Figure 21 shows the complete 206-mer sequence (SEQ ID NO:32). The 206-mer was generated by PCR . The M13/pUC 24-mer reverse sequencing (-48) primer and the M13/pUC sequencing (-47) primer from New England Biolabs (catalogue nos. 1233 and 1224 respectively) were used (50 pmoles each) with the pGEM3z(f+) plasmid vector (Promega Corp.) as template (10 ng) containing the target sequences. The conditions for PCR were as follows: 50 µM of each dNTP and 2.5 units of Taq DNA polymerase in 100 µl of 20 mM Tris-Cl, pH 8.3, 1.5 mM $MgCl_2$, 50 mM KCl with 0.05% Tween-20 and 0.05% NP-40. Reactions were cycled 35 times through 95°C for 45 seconds, 63°C for 45 seconds, then 72°C for 75 seconds. After cycling, reactions were finished off with an incubation at 72°C for 5 minutes. The resulting fragment was purified by electrophoresis through a 6% polyacrylamide gel (29:1 cross link) in a buffer of 45 mM Tris-Borate, pH 8.3, 1.4 mM EDTA, visualized by ethidium bromide staining or autoradiography, excised from the gel, eluted by passive diffusion, and concentrated by ethanol precipitation.

Cleavage of the substrate DNA occurred in the presence of the pilot oligonucleotide 19-12 at 50°C (Figure 12B, lanes 1 and 7) but not at 75°C (lanes 4 and 10). In the presence of oligonucleotide 30-12 cleavage was observed at both temperatures. Cleavage did not occur in the absence of added oligonucleotides (lanes 3, 6 and 12) or at about 80°C even though at 50°C adventitious structures in the substrate allowed primer-independent cleavage in the absence of KCl (Figure 12B, lane 9). A non-specific oligonucleotide with no complementarity to the substrate DNA did not direct cleavage at 50°C, either in the absence or presence of 50 mM KCl (lanes 13 and 14). Thus, the specificity of the cleavage reactions can be controlled by the extent of complementarity to the substrate and by the conditions of incubation.

**D. Cleavage Of RNA**

An shortened RNA version of the sequence used in the transcleavage experiments discussed above was tested for its ability to serve as a substrate in the reaction. The RNA is cleaved at the expected place, in a reaction that is dependent upon the presence of the pilot oligonucleotide. The RNA substrate, made by T7 RNA polymerase in the presence of [$\alpha$-$^{32}$P]UTP, corresponds to a truncated version of the DNA substrate used in Figure 12B. Reaction conditions were similar to those in used for the DNA substrates described above, with 50 mM KCl; incubation was for 40 minutes at 55°C. The pilot oligonucleotide used is termed 30-0 (SEQ ID NO:20) and is shown in Figure 13A.

The results of the cleavage reaction is shown in Figure 13B. The reaction was run either in the presence or absence of DNAP*Taq* or pilot oligonucleotide as indicated in Figure 13B.

Strikingly, in the case of RNA cleavage, a 3' arm is not required for the pilot oligonucleotide. It is very unlikely that this cleavage is due to previously described RNaseH, which would be expected to cut the RNA in several places along the 30 base-pair long RNA-DNA duplex. The 5' nuclease of DNAP*Taq* is a structure-specific RNaseH that cleaves the RNA at a single site near the 5' end of the heteroduplexed region.

It is surprising that an oligonucleotide lacking a 3' arm is able to act as a pilot in directing efficient cleavage of an RNA target because such oligonucleotides are unable to direct efficient cleavage of DNA targets using native DNAPs. However, some 5' nucleases of the present invention (for example, clones E, F and G of Figure 4) can cleave DNA in the absence of a 3' arm. In other words, a non-extendable cleavage structure is not required for specific cleavage with some 5' nucleases of the present invention derived from thermostable DNA polymerases.

We tested whether cleavage of an RNA template by DNAP*Taq* in the presence of a fully complementary primer could help explain why DNAP*Taq* is unable to extend a DNA oligonucleotide on an RNA template, in a reaction resembling that of reverse transcriptase. Another thermophilic DNAP, DNAPTth, is able to use RNA as a template, but only in the presence of Mn++, so we predicted that this enzyme would not cleave RNA in the presence of this cation. Accordingly, we incubated an RNA molecule with an appropriate pilot oligonucleotide in the presence of DNAP*Taq* or DNAPTth, in buffer containing either Mg++ or Mn++. As expected, both enzymes cleaved the RNA in the presence of Mg++. However, DNAP*Taq*, but not DNAPTth, degraded the RNA in the presence of Mn++. We conclude that the 5' nuclease activities of many DNAPs may contribute to their inability to use RNA as templates.

## EXAMPLE 2

Generation Of 5' Nucleases From Thermostable DNA Polymerases

Thermostable DNA polymerases were generated which have reduced synthetic activity, an activity that is an undesirable side-reaction during DNA cleavage in the detection assay of the invention, yet have maintained thermostable nuclease activity. The result is a thermostable polymerase which cleaves nucleic acids DNA with extreme specificity.

Type A DNA polymerases from eubacteria of the genus *Thermus* share extensive protein sequence identity (90% in the polymerization domain, using the Lipman-Pearson method in the DNA analysis software from DNAStar, WI) and behave similarly in both polymerization and nuclease assays. Therefore, we have used the genes for the DNA polymerase of *Thermus aquaticus* (DNAP*Taq*) and *Thermus flavus* (DNAPTfl) as representatives of this class. Polymerase genes from other eubacterial organisms, such as *Thermus thermophilus*, *Thermus sp.*, *Thermotoga maritima*, *Thermosipho africanus* and *Bacillus stearothermophilus* are equally suitable. The DNA polymerases from these thermophilic organisms are capable of surviving and performing at elevated temperatures, and can thus be used in reactions in which temperature is used as a selection against non-specific hybridization of nucleic acid strands.

The restriction sites used for deletion mutagenesis, described below, were chosen for convenience. Different sites situated with similar convenience are available in the *Thermus thermophilus* gene and can be used to make similar constructs with other Type A polymerase genes from related organisms.

### A. Creation Of 5' Nuclease Constructs

#### 1. Modified DNAP*Taq* Genes

The first step was to place a modified gene for the *Taq* DNA polymerase on a plasmid under control of an inducible promoter. The modified *Taq* polymerase gene was isolated as follows: The *Taq* DNA polymerase gene was amplified by polymerase chain reaction from genomic DNA from *Thermus aquaticus*, strain YT-1 (Lawyer *et al.*, *supra*), using as primers the oligonucleotides described in SEQ ID NOS:13-14. The resulting fragment of DNA has a recognition sequence for the restriction endonuclease EcoRI at the 5' end of the coding sequence and a BglII sequence at the 3' end. Cleavage with BglII leaves a 5' overhang or "sticky end" that is compatible with the end generated by BamHI. The PCR-amplified DNA was digested with EcoRI

and BamHI. The 2512 bp fragment containing the coding region for the polymerase gene was gel purified and then ligated into a plasmid which contains an inducible promoter.

In one embodiment of the invention, the pTTQ18 vector, which contains the hybrid *trp-lac* (*tac*) promoter, was used [M.J.R. Stark, *Gene* 5:255 (1987)] and shown in Figure 14. The *tac* promoter is under the control of the *E. coli lac* repressor. Repression allows the synthesis of the gene product to be suppressed until the desired level of bacterial growth has been achieved, at which point repression is removed by addition of a specific inducer, isopropyl-β-D-thiogalactopyranoside (IPTG). Such a system allows the expression of foreign proteins that may slow or prevent growth of transformants.

Bacterial promoters, such as *tac*, may not be adequately suppressed when they are present on a multiple copy plasmid. If a highly toxic protein is placed under control of such a promoter, the small amount of expression leaking through can be harmful to the bacteria. In another embodiment of the invention, another option for repressing synthesis of a cloned gene product was used. The non-bacterial promoter, from bacteriophage T7, found in the plasmid vector series pET-3 was used to express the cloned mutant *Taq* polymerase genes [Figure 15; Studier and Moffatt, *J. Mol. Biol.* 189:113 (1986)]. This promoter initiates transcription only by T7 RNA polymerase. In a suitable strain, such as BL21(DE3)pLYS, the gene for this RNA polymerase is carried on the bacterial genome under control of the *lac* operator. This arrangement has the advantage that expression of the multiple copy gene (on the plasmid) is completely dependent on the expression of T7 RNA polymerase, which is easily suppressed because it is present in a single copy.

For ligation into the pTTQ18 vector (Figure 14), the PCR product DNA containing the *Taq* polymerase coding region (mut*Taq*, clone 4B, SEQ ID NO:21) was digested with EcoRI and BglII and this fragment was ligated under standard "sticky end" conditions [Sambrook *et al. Molecular Cloning*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, pp. 1.63-1.69 (1989)] into the EcoRI and BamHI sites of

the plasmid vector pTTQ18. Expression of this construct yields a translational fusion product in which the first two residues of the native protein (Met-Arg) are replaced by three from the vector (Met-Asn-Ser), but the remainder of the natural protein would not change. The construct was transformed into the JM109 strain of *E. coli* and the transformants were plated under incompletely repressing conditions that do not permit growth of bacteria expressing the native protein. These plating conditions allow the isolation of genes containing pre-existing mutations, such as those that result from the infidelity of *Taq* polymerase during the amplification process.

Using this amplification/selection protocol, we isolated a clone (depicted in Figure 4B) containing a mutated *Taq* polymerase gene (mut*Taq*, clone 4B). The mutant was first detected by its phenotype, in which temperature-stable 5' nuclease activity in a crude cell extract was normal, but polymerization activity was almost absent (approximately less than 1% of wild type *Taq* polymerase activity).

DNA sequence analysis of the recombinant gene showed that it had changes in the polymerase domain resulting in two amino acid substitutions: an A to G change at nucleotide position 1394 causes a Glu to Gly change at amino acid position 465 (numbered according to the natural nucleic and amino acid sequences, SEQ ID NOS:1 and 4) and another A to G change at nucleotide position 2260 causes a Gln to Arg change at amino acid position 754. Because the Gln to Gly mutation is at a nonconserved position and because the Glu to Arg mutation alters an amino acid that is conserved in virtually all of the known Type A polymerases, this latter mutation is most likely the one responsible for curtailing the synthesis activity of this protein. The nucleotide sequence for the Figure 4B construct is given in SEQ ID NO:21. The corresponding amino acid sequence encoded by the nucleotide sequence of SEQ ID NO:21 is listed in SEQ ID NO:85.

Subsequent derivatives of DNAP*Taq* constructs were made from the mut*Taq* gene, thus, they all bear these amino acid substitutions in addition to their other alterations, unless these particular regions were deleted. These mutated sites are indicated by black boxes at these locations in the diagrams in Figure 4. In Figure 4,

the designation "3' Exo" is used to indicate the location of the 3' exonuclease activity associated with Type A polymerases which is not present in DNAP*Taq*. All constructs except the genes shown in Figures 4E, F and G were made in the pTTQ18 vector.

The cloning vector used for the genes in Figures 4E and F was from the commercially available pET-3 series, described above. Though this vector series has only a BamHI site for cloning downstream of the T7 promoter, the series contains variants that allow cloning into any of the three reading frames. For cloning of the PCR product described above, the variant called pET-3c was used (Figure 15). The vector was digested with BamHI, dephosphorylated with calf intestinal phosphatase, and the sticky ends were filled in using the Klenow fragment of DNAPEc1 and dNTPs. The gene for the mutant *Taq* DNAP shown in Figure 4B (mut*Taq*, clone 4B) was released from pTTQ18 by digestion with EcoRI and SalI, and the "sticky ends" were filled in as was done with the vector. The fragment was ligated to the vector under standard blunt-end conditions (Sambrook *et al.*, *Molecular Cloning*, *supra*), the construct was transformed into the BL21(DE3)pLYS strain of *E. coli*, and isolates were screened to identify those that were ligated with the gene in the proper orientation relative to the promoter. This construction yields another translational fusion product, in which the first two amino acids of DNAP*Taq* (Met-Arg) are replaced by 13 from the vector plus two from the PCR primer (Met-Ala-Ser-Met-Thr-Gly-Gly-Gln-Gln-Met-Gly-Arg-Ile-Asn-Ser) (SEQ ID NO:29).

Our goal was to generate enzymes that lacked the ability to synthesize DNA, but retained the ability to cleave nucleic acids with a 5' nuclease activity. The act of primed, templated synthesis of DNA is actually a coordinated series of events, so it is possible to disable DNA synthesis by disrupting one event while not affecting the others. These steps include, but are not limited to, primer recognition and binding, dNTP binding and catalysis of the inter-nucleotide phosphodiester bond. Some of the amino acids in the polymerization domain of DNAPEcI have been linked to these functions, but the precise mechanisms are as yet poorly defined.

One way of destroying the polymerizing ability of a DNA polymerase is to delete all or part of the gene segment that encodes that domain for the protein, or to otherwise render the gene incapable of making a complete polymerization domain. Individual mutant enzymes may differ from each other in stability and solubility both inside and outside cells. For instance, in contrast to the 5' nuclease domain of DNAPEcI, which can be released in an active form from the polymerization domain by gentle proteolysis [Setlow and Kornberg, *J. Biol. Chem.* 247:232 (1972)], the *Thermus* nuclease domain, when treated similarly, becomes less soluble and the cleavage activity is often lost.

Using the mutant gene shown in Figure 4B as starting material, several deletion constructs were created. All cloning technologies were standard (Sambrook *et al.*, *supra*) and are summarized briefly, as follows:

Figure 4C: The mut*Taq* construct was digested with PstI, which cuts once within the polymerase coding region, as indicated, and cuts immediately downstream of the gene in the multiple cloning site of the vector. After release of the fragment between these two sites, the vector was re-ligated, creating an 894-nucleotide deletion, and bringing into frame a stop codon 40 nucleotides downstream of the junction. The nucleotide sequence of this 5' nuclease (clone 4C) is given in SEQ ID NO:9. The corresponding amino acid sequence encoded by the nucleotide sequence of SEQ ID NO:9 is listed in SEQ ID NO:86.

Figure 4D: The mut*Taq* construct was digested with NheI, which cuts once in the gene at position 2047. The resulting four-nucleotide 5' overhanging ends were filled in, as described above, and the blunt ends were re-ligated. The resulting four-nucleotide insertion changes the reading frame and causes termination of translation ten amino acids downstream of the mutation. The nucleotide sequence of this 5' nuclease (clone 4D) is given in SEQ ID NO:10. The corresponding amino acid sequence encoded by the nucleotide sequence of SEQ ID NO:10 is listed in SEQ ID NO:87.

Figure 4E: The entire mut*Taq* gene was cut from pTTQ18 using EcoRI and SalI and cloned into pET-3c, as described above. This clone was digested with BstXI and XcmI, at unique sites that are situated as shown in Figure 4E. The DNA was treated with the Klenow fragment of DNAPEc1 and dNTPs, which resulted in the 3' overhangs of both sites being trimmed to blunt ends. These blunt ends were ligated together, resulting in an out-of-frame deletion of 1540 nucleotides. An in-frame termination codon occurs 18 triplets past the junction site. The nucleotide sequence of this 5' nuclease (clone 4E) is given in SEQ ID NO:11 [The corresponding amino acid sequence encoded by the nucleotide sequence of SEQ ID NO:11 is listed in SEQ ID NO:88]., with the appropriate leader sequence given in SEQ ID NO:30 [The corresponding amino acid sequence encoded by the nucleotide sequence of SEQ ID NO:30 is listed in SEQ ID NO:89.. It is also referred to as the enzyme Cleavase™ BX.

Figure 4F: The entire mut*Taq* gene was cut from pTTQ18 using EcoRI and SalI and cloned into pET-3c, as described above. This clone was digested with BstXI and BamHI, at unique sites that are situated as shown in the diagram. The DNA was treated with the Klenow fragment of DNAPEc1 and dNTPs, which resulted in the 3' overhang of the BstX I site being trimmed to a blunt end, while the 5' overhang of the Bam HI site was filled in to make a blunt end. These ends were ligated together, resulting in an in-frame deletion of 903 nucleotides. The nucleotide sequence of the 5' nuclease (clone 4F) is given in SEQ ID NO:12. It is also referred to as the enzyme Cleavase™ BB. The corresponding amino acid sequence encoded by the nucleotide sequence of SEQ ID NO:12 is listed in SEQ ID NO:90.

Figure 4G: This polymerase is a variant of that shown in Figure 4E. It was cloned in the plasmid vector pET-21 (Novagen). The non-bacterial promoter from bacteriophage T7, found in this vector, initiates transcription only by T7 RNA polymerase. *See* Studier and Moffatt, *supra*. In a suitable strain, such as (DES)pLYS, the gene for this RNA polymerase is carried on the bacterial genome under control of the *lac* operator. This arrangement has the advantage that expression of the multiple copy gene (on the plasmid) is completely dependent on the expression of T7 RNA

polymerase, which is easily suppressed because it is present in a single copy. Because the expression of these mutant genes is under this tightly controlled promoter, potential problems of toxicity of the expressed proteins to the host cells are less of a concern.

The pET-21 vector also features a "His-Tag", a stretch of six consecutive histidine residues that are added on the carboxy terminus of the expressed proteins. The resulting proteins can then be purified in a single step by metal chelation chromatography, using a commercially available (Novagen) column resin with immobilized $Ni^{++}$ ions. The 2.5 ml columns are reusable, and can bind up to 20 mg of the target protein under native or denaturing (guanidine-HCl or urea) conditions.

*E. coli* (DES)pLYS cells are transformed with the constructs described above using standard transformation techniques, and used to inoculate a standard growth medium (*e.g.*, Luria-Bertani broth). Production of T7 RNA polymerase is induced during log phase growth by addition of IPTG and incubated for a further 12 to 17 hours. Aliquots of culture are removed both before and after induction and the proteins are examined by SDS-PAGE. Staining with Coomassie Blue allows visualization of the foreign proteins if they account for about 3-5% of the cellular protein and do not co-migrate with any of the major host protein bands. Proteins that co-migrate with major host proteins must be expressed as more than 10% of the total protein to be seen at this stage of analysis.

Some mutant proteins are sequestered by the cells into inclusion bodies. These are granules that form in the cytoplasm when bacteria are made to express high levels of a foreign protein, and they can be purified from a crude lysate, and analyzed by SDS-PAGE to determine their protein content. If the cloned protein is found in the inclusion bodies, it must be released to assay the cleavage and polymerase activities. Different methods of solubilization may be appropriate for different proteins, and a variety of methods are known. *See e.g.*, Builder & Ogez, U.S. Patent No. 4,511,502 (1985); Olson, U.S. Patent No. 4,518,526 (1985); Olson & Pai, U.S. Patent No. 4,511,503 (1985); Jones *et al.*, U.S. Patent No. 4,512,922 (1985), all of which are hereby incorporated by reference.

The solubilized protein is then purified on the $Ni^{++}$ column as described above, following the manufacturers instructions (Novagen). The washed proteins are eluted from the column by a combination of imidazole competitor (1 M) and high salt (0.5 M NaCl), and dialyzed to exchange the buffer and to allow denatured proteins to refold. Typical recoveries result in approximately 20 μg of specific protein per ml of starting culture. The DNAP mutant is referred to as the enzyme Cleavase™ BN and the sequence is given in SEQ ID NO:31. The corresponding amino acid sequence encoded by the nucleotide sequence of SEQ ID NO:31 is listed in SEQ ID NO:91.

**2. Modified DNAPTfl Gene**

The DNA polymerase gene of *Thermus flavus* was isolated from *the "T. flavus"* AT-62 strain obtained from the American Type Tissue Collection (ATCC 33923). This strain has a different restriction map then does the *T. flavus* strain used to generate the sequence published by Akhmetzjanov and Vakhitov, *supra*. The published sequence is listed as SEQ ID NO:2. No sequence data has been published for the DNA polymerase gene from the AT-62 strain of *T. flavus*.

Genomic DNA from *T. flavus* was amplified using the same primers used to amplify the *T. aquaticus* DNA polymerase gene (SEQ ID NOS:13-14). The approximately 2500 base pair PCR fragment was digested with EcoRI and BamHI. The over-hanging ends were made blunt with the Klenow fragment of DNAPEc1 and dNTPs. The resulting approximately 1800 base pair fragment containing the coding region for the N-terminus was ligated into pET-3c, as described above. This construct, clone 5B, is depicted in Figure 5B. The wild type *T. flavus* DNA polymerase gene is depicted in Figure 5A. In Figure 5, the designation "3' Exo" is used to indicate the location of the 3' exonuclease activity associated with Type A polymerases which is not present in DNAP*Tfl*. The 5B clone has the same leader amino acids as do the DNAP*Taq* clones 4E and F which were cloned into pET-3c; it is not known precisely where translation termination occurs, but the vector has a strong transcription termination signal immediately downstream of the cloning site.